

FI Misc-difference 10 /note- "wild type Ala substd. with Val"
 FI Misc-difference 21 /note- "wild type Arg substd. with Ser"
 FI Misc-difference 23 /note- "wild type Met substd. with Ile"
 FI Misc-difference 27 /note- "wild type Leu substd. with Met"
 FI Misc-difference 33 /note- "wild type Arg substd. with Glu"
 FI Misc-difference 47 /note- "wild type Arg substd. with Glu"
 FI Misc-difference 51 /note- "wild type Asp substd. with His"
 FI Misc-difference 71 /note- "wild type Lys substd. with Thr"
 FI Misc-difference 72 /note- "wild type Leu substd. with Phe"
 FI Misc-difference 76 /note- "wild type His substd. with Arg"
 FI Misc-difference 78 /note- "wild type Pro substd. with Val"
 FI Misc-difference 93 /note- "wild type Arg substd. with Ser"
 FI Misc-difference 94 /note- "wild type His substd. with Tyr"
 FI Misc-difference 101 /note- "wild type Gly substd. with Gly"
 FI Misc-difference 110 /note- "wild type Arg substd. with Glu"
 FI Misc-difference 120 /note- "wild type Ser substd. with Thr"
 FI Misc-difference 136 /note- "wild type Arg substd. with Lys"
 FI Misc-difference 139 /note- "wild type Ala substd. with Gly"
 FI Misc-difference 141 /note- "wild type Ala substd. with Met"
 FI Misc-difference 142 /note- "wild type Arg substd. with His"
 FI Misc-difference 145 /note- "wild type Ser substd. with Arg"
 FI Misc-difference 149 /note- "wild type Asp substd. with Glu"
 FI Misc-difference 155 /note- "wild type Val substd. with Ile"
 FI Misc-difference 163 /note- "wild type Ser substd. with Ala"
 FI Misc-difference 171 /note- "wild type Pro substd. with Ser"
 FI Misc-difference 173 /note- "wild type Ala substd. with Thr"
 FI Misc-difference 174 /note- "wild type Val substd. with Leu"
 FI Misc-difference 176 /note- "wild type Ser substd. with Thr"
 FI Misc-difference 178 /note- "wild type Ser substd. with Pro"
 FI Misc-difference 187 /note- "wild type Cys substd. with Asn"
 FI Misc-difference 199 /note- "wild type Leu substd. with Met"
 FI Misc-difference 200 /note- "wild type Ala substd. with Thr"
 FI Misc-difference 201 /note- "wild type Gly substd. with Asp"
 FI Misc-difference 202 /note- "wild type Arg substd. with Lys"
 FI Misc-difference 208 /note- "wild type His substd. with Asp"
 FI Misc-difference 212 /note- "wild type Arg substd. with His"
 FI Misc-difference 224 /note- "wild type Arg substd. with His"

FI Misc-difference 233 /note- "wild type Ala substd. with Val"
 FI Misc-difference 234 /note- "wild type Glu substd. with Lys"
 FI Misc-difference 237 /note- "wild type Ser substd. with Pro"
 FI Misc-difference 245 /note- "wild type Ala substd. with Thr"
 FI Misc-difference 258 /note- "wild type Arg substd. with Glu"
 FI Misc-difference 259 /note- "wild type His substd. with Tyr"
 FI Misc-difference 260 /note- "wild type Ser substd. with Ala"
 FI Misc-difference 261 /note- "wild type Ser substd. with Leu"
 FI Misc-difference 263 /note- "wild type Ser substd. with Lys"
 FI Misc-difference 266 /note- "wild type Ala substd. with Pro"
 FI Misc-difference 269 /note- "wild type Arg substd. with Glu"
 FI Misc-difference 270 /note- "wild type Ala substd. with Arg"
 FI Misc-difference 272 /note- "wild type Val substd. with Ile"
 FI Misc-difference 275 /note- "wild type Thr substd. with Lys"
 FI Misc-difference 282 /note- "wild type Arg substd. with His"
 FI Misc-difference 287 /note- "wild type Ala substd. with Pro"
 FI Misc-difference 288 /note- "wild type Asp substd. with Glu"
 FI Misc-difference 300 /note- "wild type Gly substd. with Asp"
 FI Misc-difference 317 /note- "wild type Ser substd. with Thr"
 FI Misc-difference 328 /note- "wild type Ser substd. with Arg"
 FI Misc-difference 329 /note- "wild type Glu substd. with Asp"
 FI Misc-difference 337 /note- "wild type Arg substd. with Asn"
 FI Misc-difference 348 /note- "wild type Arg substd. with His"
 FI Misc-difference 364 /note- "wild type Arg substd. with His"
 FI Misc-difference 393 /note- "wild type Arg substd. with Glu"
 FI Misc-difference 395 /note- "wild type Met substd. with Ile"
 FI Misc-difference 399 /note- "wild type Ser substd. with Gly"
 FI Misc-difference 402 /note- "wild type Ser substd. with Glu"
 FI Misc-difference 403 /note- "wild type Ala substd. with Gly"
 FI Misc-difference 419 /note- "C-terminal addition to the wild type sequence"
 DN W09526399-A1.
 DN 05-OCT-1995.
 PD 24-MAR-1995: 9580-JP00541.
 PE 09-SEP-1994: 94JP-0216333.
 PR 25-MAR-1994: 94JP-0056271.
 XX (MARU) MARUKIN SHOU KK.
 PA Maru I, Ohta Y, Tsukada Y;
 PI WP1: 1995-351420/45.
 DX


```

UY 64 FAVYNNEL...EPKQVMEIARHG-ADPEARRGRQD-----GNWVF-ALDQE 106
DB 173 ---IqfIyKrkKwkwvswdseuqdyImydvhdhdpvnnelrrfwewytlInd 229
UY 107 G---KPIKQRYN-----VPSDFEAMARSGYALASNOEAKRILALAVANNVL 150
DB 240 qftIdvYkRkKystfdwtIthvntatKemtavacfwkndiq-----alcnylnkt 280
UY 151 RQRI...-----NPKINQEKSYPRGTHPLKSLAVPMILANLEMMLEPTTV 195
DB 281 dwkhsvidrphlhylnyasssqny-----dmaklntv 316
UY 196 HEVIACTVKEVMIIDPEIGIMKEAVIPIGEVVSEFQNLNPGHIFAM---WPMMDI 252
DB 417 qK-----Ihpr-----mbavt-----Ivdhdsq-----pqeslestvgewt 346
UY 253 AKRSKIRIOLQEOALAVVNLNLEAVMEEFSTITYELLRQGHIPROULEWQKLMWVILETL 312
DB 447 -----KPIdyalI---EreghysvfydydyqIphsypmkakIdp-----llcar 391
UY 413 VALAKSHIGATQCEKQWQWFEFVHY-----AMSH-----FADEVEGEWF 351
DB 492 qutayqat-----IhdylhmliaKwreantIhpsqalalmsdqakew 437
UY 452 GYL---NRKGVILNKKOK 368
DB 448 myzqpkakapwmltknk 456

```

RESULT 11
 ID W26684
 AC W26684:
 DI 27-MAR-1998 (first entry)
 DE Bovine lysosomal alpha-mannosidase (LAMAN) R221H mutant.
 KW LAMAN; lysosomal alpha-mannosidase; alpha-mannosidosis; cattle;
 KM diagnosis; screening; genetic test.
 OS Bos taurus.
 XX
 XX Key Location/Qualifiers
 FT Modified site 144
 FT Modified site /note- "N-glycosylated"
 FT Modified site 499
 FT Modified site /note- "N-glycosylated"
 FT Modified site 499
 FT Modified site /note- "N-glycosylated"
 FT Modified site 644
 FT Modified site /note- "N-glycosylated"
 FT Modified site 640
 FT Modified site /note- "N-glycosylated"
 FT Modified site 681
 FT Modified site /note- "N-glycosylated"
 FT Modified site 755
 FT Modified site /note- "N-glycosylated"
 FT Modified site 919
 FT Modified site /note- "N-glycosylated"
 XX
 EN W09726369-A1.
 DB 24-JUL-1997.
 XX
 XX 15-JAN-1997: 97W0-GH00109.
 XX
 FT 15-JAN-1996: 96N0-0000163.
 XX
 PA (HEB67) HEB67 T.
 PA (HEB67) HEB67 T.
 PA (NUS7) NUSSEN G.

PA (COLL/) TOLLERSKID OK.
 PA (DZIE/) DZIELEWSKA H.
 XX
 PI Berg T, Nilsen G, Tollerud OK.
 XX
 DB WP1: 1997-08-05/2/35.
 DB N-PSDB: 191095.
 XX
 PT Diagnosis and screening for bovine alpha-mannosidosis by detecting
 PT mutation(s) in alpha-mannosidase gene, also nucleic acid encoding
 PT the enzyme and derived oligonucleotide primers
 XX
 PS Example 2: Page 7: 85pp: English.
 XX
 CC This glycosylated polypeptide comprises a mutant bovine lysosomal
 CC alpha-mannosidase (LAMAN) that causes bovine alpha-mannosidosis
 CC (barn) in Galloisway. The R221H mutation results from a point mutation
 CC of the LAMAN coding sequence (see 191094). A claimed method for
 CC diagnosis or screening for barn comprises detecting the presence or
 CC absence of barn causing mutations in the LAMAN gene; another point
 CC mutation causing barn in Angus and related breeds of cattle has also
 CC been identified. PCR primers (see 191098-99) are provided for use
 CC in a claimed method for detecting these mutations. Compared with
 CC known enzymatic methods, smaller samples are needed. DNA is stable
 CC during transport and storage, and the test is more reliable and
 CC quicker.
 CC (NB: the sequence for the R221H mutant alpha-mannosidase was
 CC produced by adaptation of the native sequence (see W26682) provided
 CC in the specification).
 XX
 SO Sequence 999 AA:

Query Match 414: Score 88: DB 14: Length 999:
 Best Local Similarity 16.4%: Pred. No. 4.2:
 Matches 63: Conservative 44: Mismatches 192: Indels 144: Gaps 14:

```

UY 108 KPIKQRYNVSDFEAMARSGYALASNOEAKRILALAVANNVLKROHNPQD----- 159
DB 26 ralfprlprlssltvllaaqwaayktcpkvpkmhlnhlyphthdwwlktvdy 85
UY 160 ---YKSYV-GIPIKSIAYVMILANL-----LMEWLLPQIVIEVIAQTVK 204
DB 86 fytlyntqIgaqyylIasvssllanprtrfryolatstswqftraqktrvtr 145
UY 205 EVMTPFLDPEIGIMKEAVIPIGEVVSEFSGRL-----LNP-GH-- 241
DB 146 qrtcfanqwmvdcarthyatIdqntlrtrflocrtasdqrpvywhdptdbscg 205
UY 242 -----GIFAWPMMDIAGRSIDQIOLQALAVVNLNLEAVM-----DEHPDQI 284
DB 206 asllamtdqdlI-----qhdlyqdkkvkklImeywastsklprtdlrrsv 257
UY 285 FVYIDRQCHITQDLEWMDQ----- 302
DB 258 ---IpmynpnyqIcdmIcadvwvdrtrspynakelvyrtklardqklytktdv 314
UY 303 -----KLMWVILETLVALAKSHQ-ATG-----QEKQWQWFEFVHYVAMS- 440
DB 315 mtmgstfyenauktwtkndkIqIvnaqtrdnatrvnlystacytwolkanlswsv 474
UY 341 -----HEAPYGEWFGYTNKK 367
DB 375 KKadIIPyadqymIwtdytsst 397

```

RESULT 12
 ID W26683
 AC W26683:
 DI 27-MAR-1998 (first entry)

CC A rat cDNA clone was isolated from a spleen library in lambda-gt10
 CC on the basis of homology to the human integrin alpha-d subunit
 CC gene. RACE and PCR amplification were used to obtain the complete
 CC rat alpha-d sequence given in 991715. Rat alpha-d 1 domain/human
 CC 1934 fusions were used to raise MAbs.

XX Sequence 1161 AA:

Query Match 4.18; Score 87.5; DB 16; Length 1161;

Best Local Similarity 20.28; Pred. No. 4.5; Mismatches 123; Indels 141; Gaps 19;

Matches 79; Conservative 49; Mismatches 123; Indels 141; Gaps 19;

101 FALDQEKPLKQPYNVESEFAMAFSQAASQAQAKAIAIQVNNVLRQHNRKQY 160
 155 LLDASQSLQDIAQMKD-----LKAIMEIASLISIMYSNLI----- 198
 161 EKSPGCTRPKSLAVPMILANLHEMFWLPP-----TVEVEI-----AQVREV 206
 199-KHITLLEKLLDQSLVDPVQALTYATATITMGLFHSKQSTKSAKKILIV 257
 207 MTD-----FLDP-----ELGMEAVPTGEVDSFEGRLN-----PQHTEA 245
 258 ITDQKYPDLYSDVIPADKAGITTYATQVADATQPTAKKOLNDSAPQDHFQV 317
 246 MFMMDIAQRSGDQKQLOEQAIAV-----VINTL-EY 275
 318 QNL-----ADLSIQGLQPKLALQIGTSSSSSIQHEMSQSSALISDAPVDAQVSL 374
 276 AMDEFEGGILPYFLDKQCHP-----QQLFMQDKIMVHLELIVAAKQ----- 318
 319-----EKCWMQFERVHDYAMSHFADPEYGEWFRY-----LNKREVEVLNLR 365
 426 PRHHTGKVVITIGQATHWTRPK-----SEVATQIGSYTQASISVDVDRDQSDVIL 479
 466 G-----GKMKQCFHVP--FALMLCAETL 486
 480 QAPHYEQQLDQYQSVLPDPVAVRWPEATL 511

RESULT 14
 W24062
 ID W24062 standard: Protein: 1161 AA.
 XX
 AC W24062;
 XX
 24 FEB-1998 (first entry)
 XX
 DE Rat beta 2 integrin alpha d subunit.
 XX
 KW beta 2 integrin alpha d subunit; rat; cell migration;
 KW cell adhesion; phagocytosis; diabetes; atherosclerosis;
 KW multiple sclerosis; asthma; psoriasis; lung inflammation;
 KW acute respiratory distress syndrome; rheumatoid arthritis;
 KW hyalidoma; monoclonal antibody.
 XX
 AS Rattus sp.
 XX
 FH Key Location/Qualifiers
 FI Misc-difference 474
 FT /note- "encoded by ACY"
 FI Misc-difference 1128
 FI /note- "encoded by TYC"
 XX
 FN W09731099-A1.
 XX
 28-AUG-1997.
 XX
 24-FEB-1997; 97WD-0802713.
 XX
 22-FEB-1996; 96US-0605672.

XX (1998) 100S 009P.

PI Gallatin KM. Van der Vliet M.

XX WPI: 1997-435154/40.

DE N-PSDB: T79257.

PT Hyalidoma 199M and antibody secreted by it. Specific for non-rat
 PT beta2 integrin subunit, useful to detect subunit in cells and
 PT modulate its activity

XX Example 17: Page 170-175; 222pp; English.

CC This polypeptide comprises the rat homologue of a novel human beta
 CC 2 integrin subunit, designated alpha d (see W24047). Its sequence
 CC was deduced from a cDNA clone (see 179257) isolated from a spleen
 CC cDNA library. Alpha d is involved in cell migration, phagocytosis
 CC and cell-cell interaction. Recombinant alpha d polypeptides can be
 CC expressed in transformed host cells for use in assays for
 CC identifying antibodies or other compounds that modulate alpha d
 CC activity or which modulate the interaction between alpha d and a
 CC ligand, for treatment or preventing diseases in which macrophages are
 CC implicated. A monoclonal antibody that is specific for rat alpha d
 CC subunit, and which is secreted by hybridoma 199M (ATCC HB 12058),
 CC is claimed. It may be used to modulate alpha d activity.
 CC interfere with human alpha d, and treat or prevent diseases such as
 CC type 1 diabetes, atherosclerosis, multiple sclerosis, asthma,
 CC psoriasis, lung inflammation, acute respiratory distress syndrome
 CC and rheumatoid arthritis.

XX Sequence 1161 AA:

Query Match 4.18; Score 87.5; DB 16; Length 1161;

Best Local Similarity 20.28; Pred. No. 4.5;

Matches 79; Conservative 49; Mismatches 123; Indels 141; Gaps 19;

101 FALDQEKPLKQPYNVESEFAMAFSQAASQAQAKAIAIQVNNVLRQHNRKQY 160
 155 LLDASQSLQDIAQMKD-----LKAIMEIASLISIMYSNLI----- 198
 161 EKSPGCTRPKSLAVPMILANLHEMFWLPP-----TVEVEI-----AQVREV 206
 199-KHITLLEKLLDQSLVDPVQALTYATATITMGLFHSKQSTKSAKKILIV 257
 207 MTD-----FLDP-----ELGMEAVPTGEVDSFEGRLN-----PQHTEA 245
 258 ITDQKYPDLYSDVIPADKAGITTYATQVADATQPTAKKOLNDSAPQDHFQV 317
 246 MFMMDIAQRSGDQKQLOEQAIAV-----VINTL-EY 275
 318 QNL-----ADLSIQGLQPKLALQIGTSSSSSIQHEMSQSSALISDAPVDAQVSL 374
 276 AMDEFEGGILPYFLDKQCHP-----QQLFMQDKIMVHLELIVAAKQ----- 318
 319-----EKCWMQFERVHDYAMSHFADPEYGEWFRY-----LNKREVEVLNLR 365
 426 PRHHTGKVVITIGQATHWTRPK-----SEVATQIGSYTQASISVDVDRDQSDVIL 479
 466 G-----GKMKQCFHVP--FALMLCAETL 486
 480 QAPHYEQQLDQYQSVLPDPVAVRWPEATL 511

RESULT 15

W72824

ID W72824 standard: Protein: 1161 AA.

XX
 XX W72824;
 XX

GeneCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

10M protein - protein search, using SW model

Run on: January 1, 2001, 22:22:04 : Search time 59.08 seconds
(without alignments)
94,870 Million cell updates/sec

Hit: us-09-645-321-1

Percent score: 21.37
Sequence: 1 MAAHRKQILACVYVALHQLD.....CPHVKALMFAELLQLLYS 391

Scoring table:
Gapop 10.0 , Gapext 0.5

Search: 164575 seqs, 1676186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum E score length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0.8
Maximum Match 100%

Listing first 45 summaries

Database: Issued: Patents_AA: *
1: /cgn2_6/prodata/2/1aa/5A_00M8.pep: *
2: /cgn2_6/prodata/2/1aa/5H_00M8.pep: *
3: /cgn2_6/prodata/2/1aa/6_00M8.pep: *
4: /cgn2_6/prodata/2/1aa/PTUS_00M8.pep: *
5: /cgn2_6/prodata/2/1aa/backlist1.pep: *

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	694	42.3	402	1	US-08-553-703A-4
2	694	42.3	402	2	US-09-006-021-4
3	691	42.2	402	1	US-08-553-703A-1
4	691	42.2	402	2	US-09-006-021-1
5	668	41.1	419	1	US-08-553-703A-3
6	668	41.1	419	2	US-09-006-021-3
7	645	40.0	417	1	US-08-553-703A-2
8	645	40.0	417	2	US-09-006-021-2
9	87.5	4.1	1161	1	US-08-485-618-55
10	87.5	4.1	1161	2	US-08-485-618-55
11	87.5	4.1	1161	2	US-08-605-672-55
12	87.5	4.1	1161	2	US-08-482-293A-55
13	87.5	4.1	1161	2	US-08-943-463-55
14	86	4.0	485	2	US-08-446-803-2
15	86	4.0	485	2	US-08-861-872-2
16	86	4.0	485	3	US-08-600-656-2
17	84.5	3.9	729	1	US-08-070-165F-6
18	84.5	3.9	729	2	US-08-885-418-6
19	84.5	3.9	731	1	US-08-070-165F-10
20	84.5	3.9	731	2	US-08-885-418-10
21	81	3.8	357	2	US-08-476-254-11
22	81	3.8	359	5	5474933-9
23	79.5	3.7	1155	1	US-08-286-889-46
24	79.5	3.7	1155	1	US-08-485-618-46
25	79.5	3.7	1155	2	US-08-362-652-46
26	79.5	3.7	1155	2	US-08-605-672-46
27	79.5	3.7	1155	2	US-08-482-293A-46
28	79.5	3.7	1155	2	US-08-943-463-46

29	79.5	4.7	1161	1	US-08-485-618-53	Sequence 53, Appl
30	79.5	4.7	1161	1	US-08-462-652-53	Sequence 53, Appl
31	79.5	4.7	1161	2	US-08-605-672-53	Sequence 53, Appl
32	79.5	4.7	1161	2	US-08-482-293A-53	Sequence 53, Appl
33	79.5	4.7	1161	2	US-08-943-463-53	Sequence 53, Appl
34	78.5	3.7	509	2	US-08-557-122A-29	Sequence 29, Appl
35	78.5	3.7	1151	1	US-08-286-889-37	Sequence 37, Appl
36	78.5	3.7	1151	1	US-08-485-618-37	Sequence 37, Appl
37	78.5	4.7	1151	1	US-08-462-652-37	Sequence 37, Appl
38	78.5	4.7	1151	2	US-08-605-672-37	Sequence 37, Appl
39	78.5	3.7	1151	2	US-08-482-293A-37	Sequence 37, Appl
40	78.5	3.7	1151	2	US-08-943-463-37	Sequence 37, Appl
41	78	3.6	616	1	US-08-385-370-2	Sequence 2, Appl
42	78	3.6	616	1	US-08-385-370-4	Sequence 4, Appl
43	77.5	3.6	129	3	US-08-722-126A-10	Sequence 10, Appl
44	77.5	3.6	129	4	PCT-US95-04258-10	Sequence 10, Appl
45	76.5	3.6	773	1	US-08-019-870-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-553-703A-4
Sequence 4, Application US/08553703A
Patent No. 5795767
GENERAL INFORMATION:
APPLICANT: MARI, ISATIMI
APPLICANT: OHRA, YASUHIRO
TITLE OF INVENTION: ENZYMEASE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPILED BY: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTA Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553, 703A
FILING DATE: 30-Nov-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Altman, Daniel E.
REGISTRATION NUMBER: 34,115
REFERENCE/WORK NUMBER:
TELEPHONE: 714-760-0404
TELEFAX: 714-760-9502
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 402 amino acids
TYPE: amino acid
STRANDNESS: single
TOPOLOGY: linear
MULTIPLE TYPE: peptide
FRAGMENT TYPE: N-terminal
US-08-553-703A-4

Query Match 32.38; Score 694; Lb 1; Length 402;
Host Local Similarity 34.88; Prod. No. 730-65;
Matches 143; Conservative 77; Mismatches 153; Indels 46; Gaps 6;

02	5	PROGLAUGYUOLJID	VLPEKESYSLBQGGYSLTLEKRGVYEDHJLONOV	6	
03	8	KREPELIGANKEVQGLDLEBVMMWML	LESHBDEHSHETLEKREKRYUOLKRYMLDQYOV	6,2	
04	6,2	WGLAVAVYNLE	PKLOMLELARGHONFLARHRE	10,0	
05	6,4	MMYGLGYEKLREPELLELAKKAGHELEL	ABAVAVPEKKOAFVLELIDBQVYKVSSTF	1,2	
06	1,4	SLDPAAMVSLVAVASLADQALALQAVNVER	KECHNIKQYKESYSGTLEKSLAVP	1,7	
07	1,2	SEBTVYIMANLEMLKAVLEARQYSEAVENHMO	QVLEWVEHQSLELQYLMVAASMSAVP	1,8	
08	1,7	MLIANLEEM	EMLEPTEVEEVALVREVMIDLEPELQREKA	2,2	
09	1,8	MMLELVQGLDLEPELELAREVQGLDHW	FAKRLLEHVCQKQZQ	4,2	
10	2,2	VTHGEHVESEBQALNPSHIL	LEAMWMMQADQHSBQGLQFOAL	AVVLENTLEVMIDPE	2,8
11	2,2	VSEHPELSESLQZBQHNQ	SHALELMSLELLEKSSRSQAKLEAHVIL	LELLEPEKSSIMAD	2,8
12	2,81	PEGLPYLHPCGSHHVVQGLDLEKRLMMVHIL	LEVALAKSHQALQGLEK	WQJPEKVEBQVAMS	4,4
13	2,89	VQGLLEQLALDHPQGLLEKARKMLMMQGLAM	LEFLEMMYSESHALMLKLEQVAVETEK		4,4
14	4,1	HEALHEGEMLVLENEPEVLELNGKEKREK	LEPEVPAVIMLEAETQGLVMS	4,1	
15	4,9	QYHOPGYQDPESTLMDQKVALLEKQLEPEKKA	PEHVPYRLAMTEPMISALLS	4,9	

RESULTS 2
US 09 006 021 4
2. Specification 4, Application US/09006021

```

1  APPLICANT: MAOUI, YASUHIRO
2  APPLICANT: CHIA, YASUHIRO
3  APPLICANT: TSURUKA, YOJI
4  TITLE OF INVENTION: EPIDERMASE
5  NUMBER OF STUDENT USE: 9
6  OPERATOR/INVENTOR ADDRESS:
7  ADDRESSER: Kudoori, MATSUGI, Ohsoun & Ito
8  STREET: 6-20 Nagaoki-cho, In-ya 16th Floor
9  CITY: Nagasaki
10 STATE: JAPAN
11 COUNTRY: JAPAN
12 ZIP: 850-0000
13
14 OPERATOR/REMARKS FROM:
15 MEDIUM TYPE: Diskette
16 COMPUTER: IBM compatible
17 OPERATING SYSTEM: DOS
18 SOFTWARE: FASTING Version 1.5
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: 05/034,006, 021
21 FILING DATE:
22 CLASSIFICATION:
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: 08/554,703
25 FILING DATE: 30 NOV 1995
26 ATTORNEY/AGENT INFORMATION:
27 NAME: Attn: David L. E.
28 PERSISTENT NUMBER: 44,115
29 PERSISTENT/LOCAL NUMBER:
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: 714 760 0404
32 TELEFAX: 714 760 9502
33 FAX:
34 INFORMATION FOR SEQ ID NO: 4:
35 SEQUENCE CHARACTERISTICS:
36 LENGTH: 402 amino acids
37 TYPE: amino acid
38 SUBSEQUENCE: single
39 TOPOLGY: linear

```

MOLECULE TYPE: Peptide
EXAMINER TYPE: N terminal
US 09,006,021-4

Query Match 32.38; Score 6.94; T8 2; Length 402;
Best Local Similarity 34.88; Pred. No. 7.4e 65;
Matches 144; Conservative 77; Mismatches 153; Indels 48; Gaps 8.

[illegible]

```

RESULT      3
US-08-553 704A 1
: Sequence 1, Application US/08553703A

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

```

;      PRIOR APPLICATION DATA:

```

MEDIUM TYPE: Diskette

Best Local Similarity 20.28; Pred. No. 2.8;
Matches 79; Conservative 49; Mismatches 123; Indels 141; Gaps 19;

QY	101	FAAQBCKPLHOPVNNVSHDCEAAMAFSGOYALASQACQAKAIALAOANNVJLRHONHKGQV	160
Db	155	FLIGSINSLNDRFOVOKKD-----FKALMGFASSTLELSLMOVSNNLL-----	198
QY	161	EKSTPTGTRPLKSLAVPMILANLTLEMWLLP-----TVEEVL-----AQVREV	206
Db	199	KTHFETFEENKLLDPQSLVDPVLOLQGLIYATGCIKTYMELPLHSKNGSRKSAKKILLV	257
QY	207	MTD---FLDP-----EILMBENVTPTRFEFDSFEGRLN-----PGHTEA	245
Db	258	ITDQKRYKBDPLEYSDVIRPADKAGITIRYALGVGDVDFQEPALKELNLTIGSAPQDHFQV	317
QY	246	MMFPMMDIAORSQDQOLQHOALAV-----VINITY-FY	275
Db	318	GNF---ALNRSLOLOQEKIFALIEGTOSRNSSTFQHEMSDEBHSALTSKQPVLAGVGF	374
QY	276	AMDEEFGIIEYFLDRGHP-----OLEMDOKLMWHETLVALAKG-----	318
Db	375	SMS---GQAFIY-----PPTRIPTFIMNSQENDMDHDSIOLGSLAAVAFKKQVSHLLGA	425
QY	319	--HEATGQ-----EKWQWERVHDVAMSHFADPEYCEMRY-----LNRREBVLNLK	365
Db	426	PRHGHGTRKVVIFTEAFARIMREK-----SEVRGTOIGSYFASLGSVDYDRDJSXDLEVI	479
QY	365	G-----CKWKGCFHVP---RALIMLCALIL	386
Db	480	GAPPIYEDUTRAGVSVFPPVQVGRKQWCEATL	511

RESULT 11
 US-08-605-672-55
 Sequence 55, Application US/080605672
 Patent No. 5817515
 GENERAL INFORMATION:
 APPLICANT: Galatin, W. Michael
 APPLICANT: Van der Vliet, Monica
 TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit
 NUMBER OF SEQUENCES: 103
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borum
 STREET: 233 South Wacker Drive, 6300
 CITY: Chicago
 STATE: Illinois
 COUNTRY: United States
 ZIP: 60606-6402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/605,672
 FILING DATE:
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/173,457
 FILING DATE: 23-DEC-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/286,889
 FILING DATE: 5-AUG-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/362,652
 FILING DATE: 21-DEC-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Williams Jr., Joseph A.
 REGISTRATION NUMBER: 38,659
 REFERENCE/DOCKET NUMBER: 27866/32684
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 312-474-6300
 TELEFAX: 312-474-0448

```

: TELEX: 25-1856
: INFORMATION FOR SEQ ID NO: 55:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1161 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-605-672-55

```

Query Match	4.18	Score	87.5	DH	2	Length	1161
Best Local Similarity	20.28	Pred. No.	2.8				
Matches	79	Conservative	49	Mismatches	125	Indels	141
				Gaps	19		

```

QY 101 FALIOEKRPIKROHYNNVSDICEAAMASQVLAJSGAOFAKALIAQVANNVJRKOHKJGY 161
Db 155 FLIDGSSSINORPDAQOKRD-----FVKALMOCFASISLISIMJOYSNL----- 198
QY 161 EKSTGTRPLKSLAVPRILLANLLEMBLLP-----TVEEVL-----AQTREV 206
Db 199 -KTHFLEPERKNIIDPOSLWIPVLOGLIYVIAIGIRIMEELFHSKNSRKSAKKILIV 257
QY 207 MTD-----FLDP-----EIGIMFEATPIIGFVDSHCHLIN-----TGHGTHA 245
Db 258 ITIDQKTRDPEYSDVILPAADKAGILRYALGVDADEFTALKELNTIGSAPPQDHFRKV 317
QY 246 MMEPMIDIAORSQDPOLOEQAIAV-----VINLIPY 275
Db 318 GNF-----AALNSIQMOIOKRFPAIRHGTOSRSSSSHOMFMSYCHSSALISIMJOYICAGVSP 374
QY 276 AMDEEFSGIFYELDROGHP-----VOLEMDOKLWMIHLELIVALKG-----318
Db 375 SMS-----GQAFILY-----PINTRIPTETINSGEHWIMHDYSTIGSIAVAHWKQVHSLIIIGA 425
QY 319 ---HQATQO-----HKQKQWFFRXVHDIAWMSHAPDRYGIEMFGY-----INRGFVILNIK 365
Db 426 PRIHQTKSVYLFTQEAHMRPK-----SEVRGTOIGSYFASLQCSVDVIDRQSSXDLVL 479
QY 366 G-----GKWKQCFHVP-----RALMIGAFEL 386
Db 480 GAPHYHQIRHQGVSVFPIVIGVRGRCWQFAAIL 511

```

RESULT 12
 US-08-482-293A-55
 Sequence 55, Application US/08482293A
 Patent No. 5831029
 GENERAL INFORMATION:
 APPLICANT: Gallatin, W. Michael
 APPLICANT: Van der Vliet, Monica
 TITLE OF INVENTION: Human 2 Integrin A1p
 NUMBER OF SEQUENCES: 103
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Bourne
 STREET: 233 South Wacker Drive, 6300 Seal Tower
 CITY: Chicago
 STATE: Illinois
 COUNTRY: United States
 ZIP: 60606-6402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patcullin release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/482,293A
 FILING DATE:
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/173,497
 FILING DATE: 23-DEC-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/286,889

FILING DATE: 5 AUG 1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/462,652
 FILING DATE: 21 DEC 1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Williams Jr., Joseph A.
 REGISTRATION NUMBER: 48,659
 REFERENCE/INCKET NUMBER: 27866/42684
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 412 474 6400
 TELEFAX: 412 474 0448
 FLEX: 25 4856
 INFORMATION FOR SEQ ID NO: 55:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1161 amino acids
 TYPE: amino acid
 topology: linear
 MODULE TYPE: protein
 US 08 462 294A 55

Query Match 4.18% Score 87.5% DB 2% Length 1161

Best Local Similarity 20.28% Ident. No. 2.8%
 Matches 79% Conservative 49% Mismatches 124% Indels 141% Gaps 19%

101 FALDGRKRLKPYNYVSLKFAAMAFSGYALASAGAEKAKALALQATNNVLRKQNRKQY 160
 115 FLLDGRSGTINQKQFAGMKD FVKALMDKFASTLSELMQYNTL 198
 191 EKSYGTRPKSLAVHMLANLLEMMILLP LVEEVL 206
 199 KILPTTERKNIIDPSLAVIPVVOGLITATGTRVMEELTRKNSRKSARKILLV 257
 207 MID FLDP FGLMREAVPTGFEVSPERGLN PRGTHA 245
 258 TLDDQKRYRFLFYSDVTPAOKAGILRYATGVRAVDFPALKELNITSGAVGHVKV 417
 246 MAFMMLDAPSRGQLOEALAV VNTL EY 275
 418 GNF AAKRSTQKQKIFATFQIGSSSSSPQHMSQRCSSALSLRQVLAAGSE 474
 276 AMDEFGTLEFVLDROGHP QGLMDQKLMVHLLELVALAK 418
 475 SMS GANFLY PNTTRPTFINNSQENVIMRQSYLAVAVFWKGVHSLTGA 425
 419 HQATQD EKQWQFERVHYVAMSHFAUPFYGMPY INRGVILLNK 465
 426 PRHDTGKVVLEFQARHWRK SEVKGITQIGYFASISVVDHROGSKDVLV 479
 466 G GKRRKRTYR KALMDVLEL 486
 480 GADHYVQDQGVSVFVAVWCRWQVVAL 511

RESULT 14
 US 08 943 464 55

Sequence 55, Application US/08943463
 Patent No. 5847478
 GENERAL INFORMATION:
 APPLICANT: Gallatin, W. Michael
 APPLICANT: Van der Vliet, Monica
 TITLE OF INVENTION: No. 5847478-1 Human 2 Inverted Alpha Subunit
 NUMBER OF SEQUENCES: 114
 ADDRESS/INVENTOR ADDRESS:
 ADDRESS: Madison, of Food, Genetics, Mortal & Born
 STREET: 233 South Market Street, 6400 South Tower
 CITY: Chicago
 STATE: Illinois
 COUNTRY: United States
 Z11: 6400 6402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible

OPERATING SYSTEM: pc Dos/MS Dos
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/943,463
 FILING DATE:
 CLASSIFICATION: 540
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/174,497
 FILING DATE: 23 DEC 1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/286,889
 FILING DATE: 5 AUG 1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/462,652
 FILING DATE: 21 DEC 1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Williams Jr., Joseph A.
 REGISTRATION NUMBER: 48,659
 REFERENCE/INCKET NUMBER: 27866/42684
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 412 474 6400
 TELEFAX: 412 474 0448
 FLEX: 25 4856
 INFORMATION FOR SEQ ID NO: 55:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1161 amino acids
 TYPE: amino acid
 topology: linear
 MODULE TYPE: protein
 US 08 943 464 55

Query Match 4.18% Score 87.5% DB 2% Length 1161

Best Local Similarity 20.28% Ident. No. 2.8%
 Matches 79% Conservative 49% Mismatches 124% Indels 141% Gaps 19%

101 FALDGRKRLKPYNYVSLKFAAMAFSGYALASAGAEKAKALALQATNNVLRKQNRKQY 160
 115 FLLDGRSGTINQKQFAGMKD FVKALMDKFASTLSELMQYNTL 198
 191 EKSYGTRPKSLAVHMLANLLEMMILLP LVEEVL 206
 199 KILPTTERKNIIDPSLAVIPVVOGLITATGTRVMEELTRKNSRKSARKILLV 257
 207 MID FLDP FGLMREAVPTGFEVSPERGLN PRGTHA 245
 258 TLDDQKRYRFLFYSDVTPAOKAGILRYATGVRAVDFPALKELNITSGAVGHVKV 417
 246 MAFMMLDAPSRGQLOEALAV VNTL EY 275
 418 GNF AAKRSTQKQKIFATFQIGSSSSSPQHMSQRCSSALSLRQVLAAGSE 474
 276 AMDEFGTLEFVLDROGHP QGLMDQKLMVHLLELVALAK 418
 475 SMS GANFLY PNTTRPTFINNSQENVIMRQSYLAVAVFWKGVHSLTGA 425
 419 HQATQD EKQWQFERVHYVAMSHFAUPFYGMPY INRGVILLNK 465
 426 PRHDTGKVVLEFQARHWRK SEVKGITQIGYFASISVVDHROGSKDVLV 479
 466 G GKRRKRTYR KALMDVLEL 486
 480 GADHYVQDQGVSVFVAVWCRWQVVAL 511

RESULT 14

US 08 446 803 2
 Sequence 2, Application US/08446803
 Patent No. 5824541
 GENERAL INFORMATION:
 APPLICANT: Ostrup, Hollo
 APPLICANT: Bissard-Franzen, Henrik
 APPLICANT: Ostrgaard, Peter Ralbek

APPLICANT: Rasmussen, Michael Delberg
 APPLICANT: Van der Zee, Pia
 TITLE OF INVENTION: Alkaline Bacillus Amylase
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSSEE: No. 58245110 No. 58245110disk of No. 58245111th America
 STREET: 405 Lexington Avenue
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10174
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/446,803
 FILING DATE: 01-June-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Harrington, James J.
 REGISTRATION NUMBER: 48,711
 REFERENCE/DOCKET NUMBER: 4157,204-US
 TELEPHONE: (212) 867-0123
 TELEFAX: (212) 878-9655
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 485 amino acids
 TYPE: amino acid
 STANDARDS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-446-803-2

Query Match 4.08; Score 86; DB 2; Length 485;
 Best Local Similarity 20.0%; Pred. No. 1.1;
 Matches 87; Conservative 46; Mismatches 127; Indels 174; Gaps 24;

07 4 HRRQLAQQYVQAIHQVLPFWKYSIDRGQGYFTCLIRKQVYDIKRWLQNRQVW 63
 126 NRNDLSIDY-----LFWMLKDFPGKQNY---SDFKWHYHFLQVMDQSKQ-FQ 174
 05 64 FAVFYNRLEKRP-QWLEIARHG-ADFLARIGRQD-----GNWYF-ALDQES- 107
 175 NRIYKFRQDKAMWIVDSNGNDYLMAYADVMHFEVNEIRKRWQHWYNTNLTNDGR 234
 07 108 -----KPLRQYN-----VSDPFAAMAFSQYALASGAQFAKALAIQAYNNVLRQ 153
 126 IDAVNHKIKYFTKRLWLTIVRNATGKEMFAVAEFTKNDLG-----ALENTLNKTNN 285
 07 154 H-----NPKQYKSYDGTIRPKLSIAVPIIANITLMEMLPTLVFV 198
 126 HSYVDFVPLHYNNYASNSGNY-----DMAKLNGIVQK- 320
 07 199 LAQTVREVMFTDLPFLIGLREAVTPTGEVDSFEGKLLNFGIGLDM--WFMMDIAQR 255
 126 -----H-----MHAVT-----FVDMHDSQ-----FQESLSFQVFWF----- 448
 07 256 SDQRQDQALAVVNTLEVAWDEFGQTFEYLDROGHPHQQDLMWVHLELVA- 315
 126 -----KFLAVALL-TEGQYPSVFTDYGTITHSVPARKAKIDP-----LEAVRNF 396
 07 316 AKQHAQIQEKQWQFRRVHY-----AMSH-----FAQFQCHMDQYL 454
 126 AYGTV-----HNYFHNNITQWLRQNTLHPNSQIALHNSAQDQKHWYV 442
 07 455 -----NRGEVLLNKIKK 668
 126 443 QGNKAGVWHDTQNK 458

RESULT 15
 US-08-861-837-2
 Sequence 2, Application US/0881837
 Patent No. 5856164
 GENERAL INFORMATION:
 APPLICANT: Ottarp, Helge
 APPLICANT: Hisard-Frantzon, Henriik
 APPLICANT: Osteraard, Peter Rahbek
 APPLICANT: Rasmussen, Michael Delberg
 APPLICANT: Van der Zee, Pia
 TITLE OF INVENTION: Alkaline Bacillus Amylase
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSSEE: No. 58561640 No. 58561640disk of No. 58561641th America
 STREET: 405 Lexington Avenue
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10174
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/861,837
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/446,803
 FILING DATE: 01-June-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Harrington, James J.
 REGISTRATION NUMBER: 48,711
 REFERENCE/DOCKET NUMBER: 4157,204-US
 TELEPHONE: (212) 867-0123
 TELEFAX: (212) 878-9655
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 485 amino acids
 TYPE: amino acid
 STANDARDS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-861-837-2

Query Match 4.08; Score 86; DB 2; Length 485;
 Best Local Similarity 20.0%; Pred. No. 1.1;
 Matches 87; Conservative 48; Mismatches 127; Indels 174; Gaps 24;

07 4 HRRQLAQQYVQAIHQVLPFWKYSIDRGQGYFTCLIRKQVYDIKRWLQNRQVW 63
 126 NRNDLSIDY-----LFWMLKDFPGKQNY---SDFKWHYHFLQVMDQSKQ-FQ 174
 05 64 FAVFYNRLEKRP-QWLEIARHG-ADFLARIGRQD-----GNWYF-ALDQES- 107
 175 NRIYKFRQDKAMWIVDSNGNDYLMAYADVMHFEVNEIRKRWQHWYNTNLTNDGR 234
 07 108 -----KPLRQYN-----VSDPFAAMAFSQYALASGAQFAKALAIQAYNNVLRQ 153
 126 IDAVNHKIKYFTKRLWLTIVRNATGKEMFAVAEFTKNDLG-----ALENTLNKTNN 285
 07 235 IDAVNHKIKYFTKRLWLTIVRNATGKEMFAVAEFTKNDLG-----ALENTLNKTNN 285
 126 -----H-----MHAVT-----FVDMHDSQ-----FQESLSFQVFWF----- 448
 07 154 H-----NPKQYKSYDGTIRPKLSIAVPIIANITLMEMLPTLVFV 198
 126 HSYVDFVPLHYNNYASNSGNY-----DMAKLNGIVQK- 320
 07 199 LAQTVREVMFTDLPFLIGLREAVTPTGEVDSFEGKLLNFGIGLDM--WFMMDIAQR 255
 126 421 -----H-----MHAVT-----FVDMHDSQ-----FQESLSFQVFWF----- 448

[illegible][illegible]

RESULT	4					
AV603012/c	107nt					
DEFINITION	AV603012	517 bp	mRNA	EST	(7-Aug-2000)	
ACCESSION	AV603012	Bos taurus kidney fetus	Bos taurus cDNA clone EIK101407			
VERSION	AV603012	3, mRNA sequence.				
KEYWORDS	AV603012.1	GI:972538				
SOURCE	EST.					
ORGANISM	cow.					
	Bos taurus					
	Eukaryota; Metazoa; Chordata; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae					
	Bovidae; Bovinae; Bos.					
	1 (bases 1 to 517)					
REFERENCE	Sugimoto, Y., Hirotsune, S., Takasuga, A., Itoh, K., Jitohzu, A. and Suzuki, H.					
TITLE	bovine cDNA sequencing					
JOURNAL	Unpublished (2000)					
COMMENT	Contact: Yoshikazu Sugimoto Animal Genetics Division Shirakawa Institute of Animal Genetics Osakura, Nishiku, Nishi-Shirakawa, Fukushima 961-8061, Japan Tel: 81-248-25-6641 Fax: 81-248-25-6725 Email: kazusugi@ccoa.ocn.ne.jp Single pass sequencing. This clone was obtained from a polyA-deleted cDNA library. Location/Qualifiers 1..517 /organism="Bos taurus" /db_xref="taxon:9913" /clone="EIK101407" /clone_1fb="Bos taurus kidney fetus" /issue_type="kidney" /dev_stage="fetus" /note="Vector: pZL; Site_1: SalI; Site_2: NotI; poly A was deleted from a NotI site"					
BASE COUNT	112 a	146 c	158 g	96 t	5 others	
ORIGIN						

Matches 69; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```

UY 707 tgcctaacccgagacagcatctgaagccatctgttcatatgacatctgcccaagct 766
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 872 tgcctaacctgagcttgcgaagcttgcctggaactacacgacttccgaagca 931
UY 767 cccagcatctcatttacaagacacatctgacatgacatctgaacccctgaatag 826
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 942 aagcgaacacaccttctcctgagcagcatctcgcactgaacccgaacccgacatag 991
UY 827 cctgaat 844
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 992 cctgaat 999

```

RESULT 3
Q20383
ID Q20383 standard; DNA: 2214 BP.

XX Q20383;

XX 16-APR-1992 (first entry)

XX ADH complex structural gene (72,000 mol. wt. protein).

XX KM Alcohol dehydrogenase; acetic acid; fermentation; ss.

XX OS Acetobacter alfoacetigenes NH-24.

XX FN JF03266988-A.

XX PD 27-NOV-1991.

XX PF 26-MAR-1990; 90JP-0073440.

XX PR 26-FEB-1990; 90JP-0042301.

XX ER 26-MAR-1990; 90JP-0073440.

XX XX (NAKA-) NAKANO SUTEN KK.

XX XX WPI: 1992-019325/03.

XX DB P-PSDB: R20192.

XX XX Alcohol dehydrogenase complex structural gene - used in plasmid

XX PT and enhancing efficiency of acetic acid fermentation for

XX PI transformed acetic acid bacteria

XX PS Disclosure: Fig 4(1-3); 21pp; Japanese.

XX XX The gene encodes a protein of mol. wt. ca. 72,000. Acetobacter

XX CC transformed with the sequence can enhance the efficiency of acetic

XX CC acid fermentation. The ADH complex can be easily extracted from the

XX CC bacteria and purified and it can be used for the determination of an

XX CC alcohol.

XX CC See also Q20383-84, and -86-88.

XX CC Sequence 2214 BP; 452 A; 683 C; 662 G; 417 T; 0 other;

XX XX Query Match 2.9%; Score 33.6; DB 13; Length 2214;

XX XX Best Local Similarity 53.9%; Pred. No. 0.9;

XX XX Matches 69; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Db 992 cctgaat 999

RESULT 4

ID Q69001 standard; cDNA: 2751 BP.

XX Q69001;

XX DE 17-APR-1995 (first entry)

XX DE Product of alternative splicing of human protoadherin 4 mRNA.

XX KW Cadherin; protoadherin; cell adhesion molecule; ss.

XX XX Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 115-2160

XX PN W09414960-A.

XX PD 07-JUL-1994.

XX PF 23-DEC-1993; 93WO-0512588.

XX PR 29-DEC-1992; 92US-0998003.

XX PA (DOHE-) DOHENY EYE INST.

XX XX Suzuki S;

XX DB WPI: 1994-293849/36.

XX DB P-PSDB: R49144.

XX PT Polynucleotide sequences encoding new protoadherins - useful

XX PT for modulating natural binding and regulating activities.

XX PS Example: page 99-102; 114pp; English.

XX CC Two full length human cDNAs corresp. to the partial cDNAs of

XX CC HUMAN-42 and HUMAN-43 (Q68981,Q68982) were isolated

XX CC from human fetal brain cDNA library. Several overlapping cDNA

XX CC clones were isolated with each probe including two cDNAs which

XX CC contained the putative entire coding sequences of two novel

XX CC proteins designated protoadherin-42 (pc42) and protoadherin 43

XX CC (pc43). The DNA and deduced AA sequences of pc42 are in

XX CC Q68997/R58906, while those of pc43 are in Q68998/R58907. Sequence

XX CC analysis of various overlapping protoadherin cDNA clones revealed

XX CC that some clones cont. unique sequences at the 3' end. The

XX CC sequences forming the boundaries of the 3' end regions are

XX CC consistent with the consensus sequence of mRNA splicing, suggesting

XX CC that these clones may corresp. to alternatively spliced mRNAs.

XX CC The DNA and AA sequences of one possible product of alternative

XX CC splicing of pc42 mRNA are given in Q68999/R58911. The DNA and AA

XX CC sequences of two possible products of alternative splicing of pc43

XX CC mRNA are respectively presented in Q69000/R58912 and Q69001/R49144.

XX XX Sequence 2751 BP; 698 A; 715 C; 696 G; 642 T; 0 other;

XX XX Query Match 2.9%; Score 33.6; DB 15; Length 2751;

XX XX Best Local Similarity 45.7%; Pred. No. 1;

XX XX Matches 117; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

UY 302 ttacttgaatctgaagcaaacctctgacacccatataacatttttcgaattact 461

Db 914 ttgaagagatctggaattgaagagccccaagtggaataattactcttcggagatcaca 974

UY 362 tgcgcgcataagccttataatcttgccttgcagcgtggagcaggaagctcaaacaca 421

Db 974 accgagccgagatgagcgaactatctcccttaagaccltgaacccgagagatgataatca 1034

```

UY 422 ttgcttgcagctacatcaagctctcagccgtccagcaatcccaaggtcaatagc 481
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1b aggttcggttcggttcgagagccacacacgtctatgagatttaccatccagccaagaca 1093
UY 482 aggaattccatccaggtacacagccctccatccctgcgcgcgcgcgcgcgcgcgc 541
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1b 1094 aggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1153
UY 542 acctcagccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 557
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1b 1154 acctcagccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1169

RESULT 5
104625 5
1b 104625 standard: cDNA: 2751 bp.
XX
XX
XX 104625:
XX
XX 29 APR 1996 (first entry)
XX
XX Alternatively spliced pcd4 coding sequence #2.
XX
XX Protocadherin: pc3; pc4; pc5; human: rat; cadherin; cell adhesion; mouse;
XX ratentia; therapy; clone; frog; fruit fly; ss.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 115..2164
XX /*aa a
XX /product: protocadherin clone 43 #2
XX
XX W09600289 A1.
XX 04 JAN 1996.
XX
XX 26 JUN 1995: 9580-05808071.
XX
XX 27 JUN 1994: 9408-02681a1.
XX
XX (DDBE ) DDBENT EYE INST.
XX
XX Suzuki S:
XX
XX W01: 1996 068871/07.
XX P-PS006: R87154.
XX
XX Polyneuropathy(s) encoding human protocadherins pc3 and pc4 and rat
XX pc5. Involved in cell cell adhesion and regulation activities
XX
XX Example 3: Page 107 110: 146pp; English.
XX
XX 104624 and 104625 represent possible alternatively spliced versions of
XX the cDNA clone of the human protocadherin sequence, designated pc43.
XX These sequences were isolated after screening a human fetal brain cDNA
XX library (containing within lambdaZapII vectors) with 32P labelled
XX versions of the sequences represented by 104605 and 104606. The most
XX abundant spliced version of pc42 is represented in 104622. The
XX cytoplasmic domain of cadherin interacts with the cytoskeleton through
XX catenins and other cytoskeleton associated proteins. The cytoplasmic
XX domain is not present in all cadherins, but in those which possess it, it
XX is essential for the cadherin's adhesive function. The cadherins which do
XX not possess a cytoplasmic domain appear to function via a different
XX method from those with a cytoplasmic domain. The proteins encoded by
XX these sequences are involved in cell-cell adhesion. The proteins may
XX have regulatory functions in the cell, as well as the cell-cell adhesion
XX properties. Antibodies produced against the encoded sequences are useful
XX for modulating the binding activity of protocadherins, and can be used
XX therapeutically.
XX
XX Sequence Z751 bp: 698 A: 715 C: 696 G: 642 T: 0 other:

```

```

Query Match 2.98; Score 33.6; 108 17; Length 2751;
Post Local Similarity 45.78; Prod No. 1;
Matches 117; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

UY 402 ttgcttgcagctacacagccctccatccctgcgcgcgcgcgcgcgcgcgcgcgcgcgc 461
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1b 914 ttgcacgcggttcggttcgagagccacacgtctatgagatttaccatccagccaagaca 973
UY 462 tgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 421
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1b 974 acctcagccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1043
UY 422 ttgcttgcagctacatcaagctctcagccgtccagcaatcccaaggtcaatagc 481
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1b 1044 aggttcggttcggttcgagagccacacgtctatgagatttaccatccagccaagaca 1093
UY 482 aggaattccatccaggtacacagccctccatccctgcgcgcgcgcgcgcgcgcgcgcgc 541
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1b 1094 aggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1153
UY 542 acctcagccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 557
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1b 1154 acctcagccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1169

RESULT 6
069000 6
1b 069000 standard: cDNA: 2789 bp.
XX
XX AC 069000:
XX
XX 17 APR 1995 (first entry)
XX
XX Product of alternative splicing of human protocadherin 43 mRNA.
XX
XX cadherin; protocadherin; cell adhesion molecule; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 115-2622
XX /*aa a
XX
XX W09414960 A.
XX 07 JUL 1994.
XX
XX 23 DEC 1993: 9480-0512588.
XX
XX 29 DEC 1992: 9408-0998003.
XX
XX (DDBE ) DDBENT EYE INST.
XX
XX Suzuki S:
XX
XX W01: 1994 294849/36.
XX P-PS006: R58912.
XX
XX Polyneuropathy sequences encoding new protocadherins useful
XX for modulating natural binding and regulating activities.
XX
XX Example: Page 93 96: 114pp; English.
XX
XX Two full length human cDNAs corresp. to the partial cDNAs of
XX H0MAN 42 and H0MAN 43 (068981,068982) were isolated
XX from human fetal brain cDNA library. Several overlapping cDNA
XX clones were isolated with each probe including two cDNAs which
XX contained the putative entire coding sequences of two novel
XX proteins designated protocadherin 42 (pc42) and protocadherin 43
XX (pc43). The DNA and deduced AA sequences of pc42 are in
XX 068997/R58906, while those of pc43 are in 068998/R58907. Sequence

```


CC analysis of various overlapping protocadherin cDNA clones revealed
 CC that some clones could, unique sequences at the 3' end. The
 CC sequences forming the boundaries of the 3' end regions are
 CC consistent with the consensus sequence of mRNA splicing, suggesting
 CC that these clones may correspond to alternatively spliced mRNAs.
 CC The DNA and AA sequences of one possible product of alternative
 CC splicing of pc42 mRNA are given in G68999/G68911. The DNA and AA
 CC sequences of two possible products of alternative splicing of pc43
 CC mRNA are respectively presented in G69000/G68912 and G69001/G69144.
 CC
 CC Sequence 2789 BP: 615 A; 781 C; 761 G; 642 T; 0 other;

Query Match: 2.98; Score 33.6; DB 15; Length 2789;
 Post Local Similarity 45.78; Prod. No. 1;
 Matches 117; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

UY 402 ttgttttgaatcagaagaacacccctggtcgaacctataacgttttttcaattgt 461
 TB 1111 1111 111 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
 TB ttgaacgaatctgaatgaagaccccaagcaatgaattattactccttgaacacaca 973
 UY 462 tctgcacatgttcttgaatcattacgttgaagcagaggaagcacaagca 421
 TB 1111 11 11 111111 11 11 11 11 11 11 11 11 11 11 11 11 11
 TB 974 accagacagcagatgaagacatcttcgaccttaaaccttgaacacgaatgaataca 1033
 UY 422 ttgcctgaagcctacaaataactcctacgcctacacacacacacacacacacac 481
 TB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
 TB 1034 aggtgcagctgaacttgaagacacacacacacacacacacacacacacacacac 1093
 UY 482 aagaatcctatcgaatcctaaacccctcaacacacacacacacacacacacacac 541
 TB 1094 aggtgcagcctac 1153
 UY 542 acccagccctgaagat 557
 TB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
 TB 1154 accagcagcagac 1169

RESULT 7
 ID 104624 standard; cDNA; 2789 BP.
 AC T04624;
 BI 29-AUG-1996 (first entry)
 XX
 DE Alternatively spliced pc43 coding sequence #1.
 XX
 KW Protocadherin; pc4; pc4; pc5; human; rat; cadherin; cell adhesion; mouse;
 KW catenin; therapy; clone; frog; fruit fly; ss.
 XX
 OS Homo sapiens.
 EH Key Location/Qualifiers
 FT CDS 115..2625
 FI /*aa- a
 PI /product= protocadherin clone 43 #1
 UN W06600289 A1.
 XX
 DE 04-JAN-1996.
 XX
 KW 26-JUN-1995; G680-0508071.
 XX
 DE 27-JUN-1994; 9405-0268161.
 XX
 PA (bHE) DEMENT EYE INST.
 XX
 PI Suzuki S;
 XX
 LR WP1: 1996-068973/07.
 LR P-PSDB: K67154.
 XX

PT Polynucleotide(s) encoding human protocadherins pc4 and pc4 and rat
 PT pc5 - involved in cell-cell adhesion and regulation activities
 PT
 XX
 XX Example 3: Page 109-104; 14pp; English.

CC T03624 and T04625 represent possible alternatively spliced versions of
 CC the cDNA clone of the human protocadherin sequence, designated pc43.
 CC These sequences were isolated after sequencing a human fetal brain cDNA
 CC library (contained within lambdaZapII vectors), with 3'p labeled
 CC versions of the sequences represented by 103605 and 103606. The most
 CC abundant spliced version of pc42 is represented in 103622. The most
 CC cytoplasmic domain of cadherin interacts with the cytoskeleton through
 CC catenins and other cytoskeleton associated proteins. The cytoplasmic
 CC domain is not present in all cadherins, but in those which possess it, it
 CC is essential for the cadherins adhesive function. The cadherins which do
 CC not possess a cytoplasmic domain appear to function via a different
 CC method from those with a cytoplasmic domain. The proteins encoded by
 CC these sequences are involved in cell-cell adhesion. The proteins may
 CC have regulatory functions in the cell, as well as the cell-cell adhesive
 CC properties. Antibodies produced against the encoded sequences are useful
 CC for modulating the binding activity of protocadherins, and can be used
 CC therapeutically.

SQ Sequence 2789 BP: 615 A; 781 C; 761 G; 642 T; 0 other;

Query Match: 2.98; Score 33.6; DB 17; Length 2789;
 Post Local Similarity 45.78; Prod. No. 1;
 Matches 117; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

UY 402 ttgttttgaatcagaagaacacccctggtcgaacctataacgttttttcaattgt 461
 TB 1111 1111 111 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
 TB ttgaacgaatctgaatgaagaccccaagcaatgaattattactccttgaacacaca 973
 UY 462 tctgcacatgttcttgaatcattacgttgaagcagaggaagcacaagca 421
 TB 1111 11 11 111111 11 11 11 11 11 11 11 11 11 11 11 11 11
 TB 974 accagacagcagatgaagacatcttcgaccttaaaccttgaacacgaatgaataca 1033
 UY 422 ttgcctgaagcctacaaataactcctacgcctacacacacacacacacacacac 481
 TB 1034 aggtgcagctgaacttgaagacacacacacacacacacacacacacacacacac 1093
 UY 482 aagaatcctatcgaatcctaaacccctcaacacacacacacacacacacacacac 541
 TB 1094 aggtgcagcctac 1153
 UY 542 acccagccctgaagat 557
 TB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
 TB 1154 accagcagcagac 1169

RESULT 8
 ID G68998 standard; cDNA; 4705 BP.
 AC G68998;
 BI 17-APR-1995 (first entry)
 XX
 DE Human protocadherin-43.
 XX
 KW cadherin; protocadherin; cell adhesion molecule; ss.
 XX
 OS Homo sapiens.
 EH Key Location/Qualifiers
 FT CDS 115..2827
 FI /*aa a
 PI W09414960-A.
 XX
 LR 07-JUL-1994.
 XX

[illegible]

DB	856	ayta	859
RESULT 11			
ID	T58681/c		
	T58681 standard; DNA: 7573 BP.		
XX			
AC	158681:		
XX			
DT	05-JAN-1998 (first entry)		
XX			
DE	DNA encoding human IGFAP1.		
XX			
KW	IGFAP1: human GTPase-activating protein; Ig motif; diagnosis; treatment;		
XX	tumour; suppressor; ras; cancer; p21-ras; neoplastic cell; ds.		
XX			
OS	Homo sapiens.		
XX			
FI	Key	Location/Challiers	
F1	CDS	468..5441	
F1		/*Tag a	
F1		/product: IGFAP1	
XX			
FN	U55639651 A.		
XX			
PD	17-JUN-1997.		
XX			
PF	09-AUG-1994: 94US-0287959.		
XX			
PR	09-AUG-1994: 94US-0287959.		
XX			
TA	(GENE) GEN HOSPITAL CORP.		
XX			
PI	Hernards A, Settleman J, Weissbach L;		
XX			
DR	WPI: 1997-332049/30.		
XX	F-PSDB: W18822.		
PI	DNA encoding GTPase-activating protein IGFAP1 - for production		
XX	recombinant protein useful for tumour diagnosis and therapy		
PS	Claim 2: Column 21-28; 35pp; English.		
XX			
CC	This sequence encodes a human GTPase-activating protein (IGFAP1). The		
CC	protein has an "IQ motif" which is defined as an amino acid sequence of		
CC	20-40 amino acids in length containing an isoleucine residue immediately		
CC	followed by a glutamine residue which has at least 50 percent sequence		
CC	similarity to the consensus sequence shown in W18822. The DNA sequence is		
CC	used for production of recombinant IGFAP1, which is useful in the		
CC	diagnosis and treatment of tumours characterized by aberrant ras		
CC	expression. Detection of mutations in the IGFAP1 gene is diagnostic of		
CC	cancer. The IGFAP1 protein can be used for treatment of cancer to reduce		
CC	the activity of p21-ras. Detection of neoplastic cells can be achieved by		
XX	measuring IGFAP1 expression.		
SO	Sequence 7573 BP; 2320 A; 1570 C; 1680 G; 2003 T; 0 other;		
Query Match	2.8%; Score 32.8; Eff 16; Length 7573;		
Best Local Similarity	49.4%; Fred. No. 3.4;		
Matches	85; Conservative 0; Mismatches 67; Indels 0; Gaps 0;		
QY	852 attatattctctgagtcgacgagcaccctcccaacacatgagatgagacaaagct 911		
DB	1872 ATTGTAATATACCTGTTGATCAGAGGTCAGCGATGACAACTTCTGTAACAGAGAGTT 1813		
QY	912 ctatgagatagacattgaaacacctgattgacctagacaaagacacacaaagacacag 971		
DB	1812 AGGAGAGGCGGAGCATTAATCTTCAGAGACTTGTGCGCTGTAGCGTACGCAAGTCTTCTGAT 1753		
QY	972 aagaaatattatgacatgatttgaagagatgacatgattgacatgacatgacat 1023		

Search completed: January 1, 2001, 22:23:45
Job Time: 12:56:58

[illegible]

```

01 1117 cccatccccacccctcctatgctctcgaggaacct 1157
02 |11111111|11|11|11|11|11|11|11|11|11|11|
03 1143 cttactgttcgcggactgcttgcctatttgctcaaacaaagtct 1181

RESULT
US 08-553-704A-9
Sequence 9, Application US/0855-6703A
Patent No. 5,955,677

GENERAL INFORMATION:
APPLICANT: MARU, ISAHUMI
APPLICANT: OHYA, YASUHIRO
APPLICANT: ISUKADA, YOJI
TITLE OF INVENTION: EPIMERASE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knolly Martens, Olson & Beal
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPILED BY: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553-704A
FILING DATE: 30-Nov-1995
CLASSIFICATION: A45
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Altman, Daniel E
REGISTRATION NUMBER: 34,115
REFERENCE/CHECK NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 714-760-0404
TELEFAX: 714-760-9502
TELEX:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1209 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1..1206
OTHER INFORMATION:
INS 08-553-704A-9

Query Match 11-48% Score 144-2; Ins 2; Length 1209;
Best Local Similarity 48-48; Pred No. 19-44;
Matches 540; Conservative 0; Mismatches 508; Indels 21; Gaps 9;

04 60 cttatgcccttttqqqaanaalattcccatgcacgcccaggagcccttatcctta 119
05 |11|11|11|11|11|11|11|11|11|11|11|11|11|11|11|11|11|11|11|
06 63 ctctatttcttcttcttgttataatcttcgacgctgagagagagagagagagagag 122
07 120 cttatagccctaaaagccaggtttttttaaccataaatcttatctaaagaacatca 179
08 |11|11|11|11|11|11|11|11|11|11|11|11|11|11|11|11|11|11|11|
09 123 ctctaaagccagagagagagagagagagagagagagagagagagagagagagagag 182
10 gtagagagagagagagagagagagagagagagagagagagagagagagagagagagag 243
11 |11|11|11|11|11|11|11|11|11|11|11|11|11|11|11|11|11|11|11|
12 183 gattttgagagcttactgtgctcttcttactgcgaagcttgcagcttttccaccttcc 242

```



```

07 702 gggatgttgcacccaaacacgacgtatgaacccatgtgtatcatgagatgccc 761
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
08 743 gaggatgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 782
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
09 762 agcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 808
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10 783 cccgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 842
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
11 819 ggaatagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 878
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12 843 cgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 902
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13 879 cgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 938
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
14 903 cgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 962
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
15 949 cgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 998
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
16 963 cgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1022
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
17 999 ggcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1058
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
18 1023 ggcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1082
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
19 1059 cgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1118
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
20 1083 cgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
21 1119 cgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1157
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
22 1143 cgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 5

US-07-985-458-1

```

1 Sequence 1, Application US/07985458
2 Patent No. 5443777
3 GENERAL INFORMATION:
4 APPLICANT: Tamaki, Toshimi;
5 APPLICANT: Takemura, Hiroshi;
6 APPLICANT: Takama, Kenji;
7 APPLICANT: Fukaya, Masahiro;
8 APPLICANT: Okumura, Hajime and
9 APPLICANT: Kawamura, Yoshiya
10 TITLE OF INVENTION: Structural gene of Membrane Bound
11 TITLE OF INVENTION: Alcohol Dehydrogenase Complex, Plasmid
12 TITLE OF INVENTION: Containing The Same And Transformed Acetic Acid
13 NUMBER OF SEQUENCES: 4
14 CORRESPONDENCE ADDRESS:
15 ADDRESSEE: Fishbait, Holtz, Goodman & Woodward, P.C.
16 STREET: 600 Third Avenue
17 CITY: New York
18 STATE: New York
19 COUNTRY: USA
20 ZIP: 10016-2088
21 COMPUTER READABLE FORM:
22 MEDIUM TYPE: 3.5 inch, 0.72 mb
23 COMPUTER: IBM PC compatible (NEC PC-9801 ES)
24 OPERATING SYSTEM: MS-DOS
25 SOFTWARE: ASCII Form
26 CURRENT APPLICATION DATA:
27 APPLICATION NUMBER: US/07/985,458
28 FILING DATE: 19921203
29 CLASSIFICATION: 435
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: US 07/658,221
32 FILING DATE: 20 FEB-1991
33 APPLICATION NUMBER: 73440/1990
34 FILING DATE: 26-MAR-1990
35 ATTORNEY/AGENT INFORMATION:
36 NAME: Goodman, Herbert
37 REGISTRATION NUMBER: 17081

```

```

1 REFERENCE/SCREEN NUMBER: 910134/80
2 TELECOMMUNICATION INFORMATION:
3 TELEPHONE: (212)972-1400
4 TELEFAX: (212) 470-1622
5 TELEX: 246268
6 INFORMATION FOR SEQ ID NO: 1:
7 SEQUENCE CHARACTERISTICS:
8 LENGTH: 2214 base pairs
9 TYPE: nucleic acid
10 STRANDEDNESS: double
11 TOPOLGY: unknown
12 MOLECULE TYPE: genomic DNA
13 ORIGINAL SOURCE:
14 ORGANISM: Acetobacter alioactiologues
15 STRAIN: MH 24
16 PUBLICATION INFORMATION:
17 AUTHORS: Tamaki, Toshimi;
18 AUTHORS: Fukaya, Masahiro;
19 AUTHORS: Takemura, Hiroshi;
20 AUTHORS: Takama, Kenji;
21 AUTHORS: Okumura, Hajime;
22 AUTHORS: Nishiyama, Makoto;
23 AUTHORS: Horinouchi, Sachiro and
24 AUTHORS: Hoppu, Teruhiko
25 TITLE: Cloning and Sequencing of the Gene Cluster
26 TITLE: Encoding Two Subunits of Membrane-Bound
27 TITLE: Alcohol Dehydrogenase from Acetobacter
28 TITLE: Polyoxydones
29 JOURNAL: Hirochimica et Biophysica Acta.
30 VOLUME: 1088
31 PAGES: 292-300
32 DATE: 1991
33 US-07-985-458-1

```

Query Match

2.9% Score 33.67 DB 1: Length 2214:

Best Local Similarity 53.9% Prod. No. 0.4:

Matches 69: Conservative 0: Mismatches 59: Indels 0: Gaps 0:

```

07 707 tgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 766
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
08 872 tgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 941
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
09 767 cgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 826
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10 932 aggcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 991
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
11 827 cgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 834
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12 992 tgcgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 999
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 6

US-07-998-003A-105

Sequence 106, Application US/07998003A

Patent No. 5643741

GENERAL INFORMATION:

APPLICANT: Suzuki, Shintaro

TITLE OF INVENTION: Protophythin Materials and Methods

NUMBER OF SEQUENCES: 107

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &

STREET: 20 South Clark Street

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60603

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

RESULT 1
 US OR 454 2748 106
 Sequence 106, Application US/08454274B
 Patent No. 5663300
 GENERAL INFORMATION:
 APPLICANT: Suzuki, Shinaro
 TITLE OF INVENTION: Polycrystalline Materials and Methods
 NUMBER OF SEQUENCES: 107
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Matsuda, O'Toolo, Costello, Murray & Hornum
 STREET: 6400 South Tower, 244 South Wacker Drive
 CITY: Chicago
 STATE: Illinois
 COUNTRY: United States of America
 ZITE: 60606-6402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT ALLOCATION DATA:
 APPLICATION NUMBER: US/08454274B

RESIDUAL 3
US-08 454 695A 106
SEQUENCE 106, APPLICATION US/0845695A
PATENT NO. 5708144
GENERAL INFORMATION:
APPLICANT: SUZUKI, Shin-aro
TITLE OF INVENTION: Protocoladder in Materials and Methods
NUMBER OF SEQUENCES: 115
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstlein, Mulrady, &
ADDRESSEE: Borum
STREET: 243 South Wacker, 6400 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC DOS/MS DOS
SOFTWARE: Patent in process #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/454,695A
FILING DATE:
CLASSIFICATION: G40
ATTORNEY/AGENT INFORMATION:


```

REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 32149
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 106:
SEQUENCE CHARACTERISTICS:
LENGTH: 2751 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 115..2160
PCT-US95-08071 106

```

```

Query Match 2.98; Score 33.6; DB 6; Length 2751;
Best Local Similarity 45.78; Pred. No. 0.45;
Matches 117; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```

```

DB 302 ttgcttgatgaggaagaaacccctgctgaacctataacgttttttccgaattc 361
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 914 ttgcaatgattctgattgaaagaccccaacgctgaaatttttttccgaattc 973
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 362 tccgcacagagctttagcaatataccttagccatggagcgaggaagctaaagca 421
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 974 accgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1093
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 422 ttgcttgatgaggaagaaacccctgctgaacctataacgttttttccgaattc 481
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1034 agggcggtaattccgaagacgacgacgacgacgacgacgacgacgacgacgacgac 1093
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 482 agagagctctatgagatataacacctgaatccctgaggtgagatgattacga 541
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1094 agggcggtaattccgaagacgacgacgacgacgacgacgacgacgacgacgacgac 1153
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 542 acctcaccctgaat 557
||| ||| ||| ||| |||
DB 1154 acgaacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1169
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 13

```

US-07-998-003A-104
Sequence 104, Application US/07998003A
Patent No. 564781
GENERAL INFORMATION:
APPLICANT: SUZUKI, Shin-aro
TITLE OF INVENTION: Protocadherin Materials and Methods
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
ADDRESS: Rockwell
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60604
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/998,003A
FILING DATE:
ATTORNEY/AGENT INFORMATION:
CLASSIFICATION: 435
NAME: No. 5643781and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 30903

```

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 2789 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 115..2622
US-07-998-003A-104

```

```

Query Match 2.98; Score 33.6; DB 1; Length 2789;
Best Local Similarity 45.78; Pred. No. 0.45;
Matches 117; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```

```

DB 302 ttgcttgatgaggaagaaacccctgctgaacctataacgttttttccgaattc 361
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 914 ttgcaatgattctgattgaaagaccccaacgctgaaatttttttccgaattc 973
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 362 tccgcacagagctttagcaatataccttagccatggagcgaggaagctaaagca 421
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 974 accgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1093
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 422 ttgcttgatgaggaagaaacccctgctgaacctataacgttttttccgaattc 481
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1034 agggcggtaattccgaagacgacgacgacgacgacgacgacgacgacgacgacgac 1093
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 482 agagagctctatgagatataacacctgaatccctgaggtgagatgattacga 541
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1094 agggcggtaattccgaagacgacgacgacgacgacgacgacgacgacgacgacgac 1153
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 542 acctcaccctgaat 557
||| ||| ||| ||| |||
DB 1154 acgaacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1169
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 14

```

US-08-453-274B-104
Sequence 104, Application US/08453274B
Patent No. 5663309
GENERAL INFORMATION:
APPLICANT: SUZUKI, Shin-aro
TITLE OF INVENTION: Protocadherin Materials and Methods
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borum
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/453,274B
FILING DATE: 30-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: No. 5663309and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 32660
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: January 1, 2001, 22:51:35 ; Search time 76.08 Seconds

(without alignments)
104.191 Million cell updates/sec

Title: US-09-645-321-1

Percent score: 2147
Sequence: 1 MIAHRPQELAQYQVALHDD.....CFHPPALMLCAETQLPVS 391

Scoring table: HUSUM62
Gapop 10.0 , Gapext 0.5

Searched: 87994 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	694	42.3	402	1 RHP_PIG	P17560 sus scrofa
2	668	41.1	419	1 RHP_RAT	P51607 ratius norv
3	645	40.0	417	1 RHP_HUMAN	P51607 homo sapien
4	201	9.4	387	1 RHP_HUMAN	P29594 rhizobium m
5	148	6.9	413	1 YHIS_ECOLI	P32140 escherichia
6	98.5	4.6	689	1 YVAL_HAUSU	P37512 bacillus su
7	97.5	4.5	1007	1 MA2H_HUMAN	P06432 felis silve
8	96	4.5	657	1 RHY_PALVO	P08088 palinurus v
9	95	4.4	876	1 SYV_MYCTO	P03175 mycobacteri
10	94.5	4.4	1010	1 MA2B_HUMAN	P00754 homo sapien
11	93.5	4.4	2493	1 CYAA_USJMA	P49606 ustilago ma
12	93	4.3	657	1 HCYR_PANIN	P10787 pennitrus i
13	92.5	4.3	748	1 GUNC_HSIPI	P27033 pseudomonas
14	91	4.2	546	1 SYL_PERO	P46214 pseudomonas
15	91	4.2	657	1 HCYA_PANIN	P40254 panulirus i
16	90.5	4.2	582	1 HEMO_OPSA	P41090 opsanus lau
17	89.5	4.2	698	1 TRPA_HUMAN	P04380 homo sapien
18	88.5	4.1	598	1 GSGL_YEAST	P06944 saccharomyc
19	88	4.1	900	1 SYA_MYCPN	P75364 mycoplasma
20	87.5	4.1	880	1 GGN4_THERP	P26221 thermomonos
21	87	4.1	999	1 MA2B_BOVIN	P24451 bos taurus
22	86	4.0	1066	1 SYL_PERO	P06792 pyrococcus
23	85.5	4.0	685	1 AMYL_DICHT	P03961 dictyoglomu
24	85.5	4.0	2405	1 DYHA_CHLHE	P36610 chlamydomon
25	85	4.0	339	1 LYCA_MORCI	P15057 bacterioph
26	85	4.0	700	1 CAN2_CHICK	P92178 gallus gall
27	85	4.0	833	1 CM41_YEAST	P51008 saccharomyc
28	84.5	3.9	570	1 RSL_CHLMT	P38016 chlamydia p
29	84.5	3.9	580	1 RSL_CHLMT	P09803 chlamydia p
30	84	3.9	758	1 PARC_RHIME	P09749 rhizobium m
31	84	3.9	2222	1 DPOF_YEAST	P21951 saccharomyc
32	83.5	4.9	470	1 NRAM_IAMIM	P05803 influenza a
33	83.5	4.9	472	1 PNB_ECOLI	P13685 escherichia

34	83.5	3.9	513	1 AR09_YEAST	P38840 saccharomyc
35	83.5	3.9	906	1 RPOD_BKRL	P19147 bacterioph
36	83.5	3.9	1839	1 CYAA_SACKL	P23466 saccharomyc
37	83	3.9	293	1 RSO_NEUCR	P01291 neorospira
38	83	3.9	578	1 TRPA_HUMAN	P19813 cytochrome
39	83	3.9	649	1 GUND_CLOTH	P04954 cytochrome
40	83	3.9	812	1 CHEA_TREPA	P96123 treponema p
41	83	3.9	884	1 RPOD_BPT3	P07459 bacterioph
42	83	3.9	901	1 MALI_ECOLI	P06993 escherichia
43	82.5	3.8	301	1 LEGL_MOUSE	P24721 mus musculu
44	82.5	3.8	489	1 TCRP_VIRCH	P24481 virus to chlo
45	82.5	3.8	528	1 MDLC_PSEPO	P20906 pseudomonas

ALIGNMENT

RESULT	ID	RHP_PIG	STANDARD	PRT	402 AA
1	AC	P17560			
DT	01-AUG-1990	(Ref. 15, Created)			
DT	01-AUG-1990	(Ref. 15, Last sequence update)			
DT	30-MAY-2000	(Ref. 39, Last annotation update)			
DE	N-ACETYLGLUCOSAMINE 2-EPIMERASE (EC 5.1.3.8) (GLUCNAZ 2-EPIMERASE)				
DE	(N-ACETYL-D-GLUCOSAMINE 2-EPIMERASE) (RENIN-BINDING PROTEIN) (RHP)				
GN	RHP				
OS	Sus scrofa (pig)				
OC	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:				
OC	Mammalia: Eutheria: Cetartiodactyla: Suidae: Sus:				
RN	[1]				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 4-12 AND 326-357.				
RC	TISSUE-KIDNEY:				
KX	MEDLINE: 90216671.				
KA	Inoue H., Fukui K., Takahashi S., Miyake Y.:				
RT	"Molecular cloning and sequence analysis of a cDNA encoding a porcine				
RT	kidney renin-binding protein."				
RL	J. Biol. Chem. 265:6556-6561(1990).				
RN	[2]				
RP	LEUCINE-ZIPPER, AND MOTACININIS.				
KX	MEDLINE: 91268065.				
KA	Inoue H., Takahashi S., Fukui K., Miyake Y.:				
RT	"Leucine zipper motif in porcine renin-binding protein (Rbp) and its				
RT	relationship to the formation of an Rbp-renal heterodimer and an				
RT	Rbp homodimer."				
RL	J. Biol. Chem. 266:11896-11900(1991).				
RN	[3]				
RP	CHARACTERIZATION.				
RC	TISSUE-KIDNEY CORTEX:				
KX	MEDLINE: 96279179.				
KA	Maru I., Ohta Y., Murata K., Tsukada Y.:				
RT	"Molecular cloning and identification of N-acetyl-D-glucosamine				
RT	2-epimerase from porcine kidney as a renin-binding protein."				
RL	J. Biol. Chem. 271:16294-16299(1996).				
CC	-1- FUNCTION: CATALYZES THE INTERCONVERSION OF N-ACETYLGLUCOSAMINE TO				
CC	N-ACETYLGLUCOSAMINE. BINDS TO RENIN FORMING A PROTEIN COMPLEX				
CC	CALLED HIGH MOLECULAR WEIGHT (HMW) RENIN AND INHIBITS RENIN				
CC	ACTIVITY.				
CC	-1- CATALYTIC ACTIVITY: N-ACETYL-D-GLUCOSAMINE - N-ACETYL-D-				
CC	GLUCNAZ 2-EPIMERASE ACTIVITY.				
CC	-1- PATHWAY: N-ACETYLFRUCTAMINIC ACID BIOSYNTHESIS.				
CC	-1- SUBUNIT: HOMODIMER.				
CC	-1- TISSUE SPECIFICITY: KIDNEY, LIVER, ADRENAL, AND PITUITARY GLANDS				
CC	THE AMOUNT BEING MUCH GREATER IN KIDNEY THAN IN THE OTHER TISSUES.				
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION				
CC	BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION				
CC	AT THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS				
CC	USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY				
CC	MOLLIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL				
CC	ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/anonymous/				

00 Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
 01
 02
 03
 04
 05
 06
 07
 08
 09
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795
 796
 797
 798
 799
 800
 801
 802
 803
 804
 805
 806
 807
 808
 809
 810
 811
 812
 813
 814
 815
 816
 817
 818
 819
 820
 821
 822
 823
 824
 825
 826
 827
 828
 829
 830
 831
 832
 833
 834
 835
 836
 837
 838
 839
 840
 841
 842
 843
 844
 845
 846
 847
 848
 849
 850
 851
 852
 853
 854
 855
 856
 857
 858
 859
 860
 861
 862
 863
 864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899
 900
 901
 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917
 918
 919
 920
 921
 922
 923
 924
 925
 926
 927
 928
 929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942
 943
 944
 945
 946
 947
 948
 949
 950
 951
 952
 953
 954
 955
 956
 957
 958
 959
 960
 961
 962
 963
 964
 965
 966
 967
 968
 969
 970
 971
 972
 973
 974
 975
 976
 977
 978
 979
 980
 981
 982
 983
 984
 985
 986
 987
 988
 989
 990
 991
 992
 993
 994
 995
 996
 997
 998
 999
 1000

Query Match
 Host Local Similarity 34.78; Score 645; DB 1; Length 417;
 Matches 148; Conservative 75; Mismatches 164; Indels 22; Gaps 7;
 01
 02
 03
 04
 05
 06
 07
 08
 09
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536

RN [3]
 RP SEQUENCE OF 23-1010 FROM N.A.
 RC TISSUE: SPLEEN;
 RA MEDLINE: 97067056.
 RA Iino Y., F., Iai A., Motemen K.W.;
 RT "Cloning, expression, purification, and characterization of the human
 RT broad specificity lysosomal acid alpha-mannosidase.";
 RL J. Biol. Chem. 271:28348-28358(1996).
 RN [4]
 RP SEQUENCE OF 23-1010 FROM N.A.
 RA MEDLINE: 9422092.
 RA Nebes V.L., Schmidt M.C.;
 RT "Human lysosomal alpha-mannosidase: isolation and nucleotide sequence
 RT of the full-length cDNA.";
 RL Biochem. Biophys. Res. Commun. 200:239-245(1994).
 RN [5]
 RP PARTIAL SEQUENCE;
 RA MEDLINE: 95144211.
 RA Enliad C., Mattioli S., Stirling J.L., Maras B., Orlandio A.;
 RT "Partial sequence of the purified protein confirms the identity of
 RL the cDNA coding for human lysosomal alpha-mannosidase B.";
 RL Biochem. J. 305:363-366(1995).
 RN [6]
 RP VARIANTS AM LEU-71: ARG-355 AND TRP-749.
 RA MEDLINE: 98431785.
 RA Goida Y., Wakamatsu N., Kawai H., Nishida Y., Matsumoto T.;
 RT "Missense and nonsense mutations in the lysosomal alpha-mannosidase
 RL gene (MANB) in severe and mild forms of alpha-mannosidosis.";
 RL Am. J. Hum. Genet. 63:1015-1024(1998).
 RN [7]
 RP VARIANTS AM, AND VARIANTS.
 RA MEDLINE: 99115084.
 RA Berg L., Risse H.M.F., Hansen G.M., Malm D., Tranobjaerg L.;
 RA Jørgensen O.K., Nilsen O.;
 RT "Spectrum of mutations in alpha-mannosidosis.";
 RL Am. J. Hum. Genet. 64:77-88(1999).
 RN [8]
 RP FUNCTION: NECESSARY FOR THE CATABOLISM OF N-LINKED CARBOHYDRATES
 RELEASED DURING GLYCOPROTEIN TURNOVER. CLEAVES ALL KNOWN TYPES OF
 ALPHA-MANNOSIDIC LINKAGES.
 RN [9]
 RP CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING
 ALPHA-D-MANNOSE RESIDUES IN ALPHA-D-MANNOSIDES.
 RN [10]
 RP SUBCELLULAR LOCATION: LYSOSOMAL.
 RN [11]
 RP PTM: FIRST PROCESSED INTO 3 PEPTIDES OF 70 KDA, 42 KDA (D) AND
 14/15 KDA (E). THE 70 KDA PEPTIDE IS FURTHER PROCESSED INTO THREE
 PEPTIDES (A, B AND C). THE A, B AND C PEPTIDES ARE
 DISULFIDE-LINKED.
 RN [12]
 RP PTM: HEAVILY GLYCOSYLATED.
 RN [13]
 RP DISEASE: DEFECTS IN MANB ARE THE CAUSE OF LYSOSOMAL ALPHA-
 MANNOSIDOSIS (AM), A LYSOSOMAL STORAGE DISEASE CHARACTERIZED BY
 ACCUMULATION OF UNBRANCHED OLIGOSACCHARIDES CHAINS. THIS
 VACUOLATION PREDOMINANTLY IN THE CNS AND PARENCHYMATOUS ORGANS,
 DEPENDING ON THE CLINICAL FINDINGS AT THE AGE OF ONSET, A SEVERE
 INFANTILE (TYPE I) AND A MILD JUVENILE (TYPE II) FORM OF ALPHA-
 MANNOSIDOSIS ARE RECOGNIZED. THERE IS CONSIDERABLE VARIATION IN
 THE CLINICAL EXPRESSION WITH MENTAL RETARDATION, RECURRENT
 INFECTIONS, IMPAIRED HEARING AND HURLER-LIKE SKELETAL CHANGES
 BEING THE MOST CONSISTENT ABNORMALITIES.
 RN [14]
 RP SIMILARITY: BELONGS TO FAMILY 38 OF GLYCOSYL HYDROLASES.
 RN [15]
 RP THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
 BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
 THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
 USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
 MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
 ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 DR EMBL: 060887; AAC51362.1;
 DR EMBL: 060899; AAC51362.1;
 DR EMBL: 060885; AAC51362.1; JOINED.
 DR EMBL: 060886; AAC51362.1; JOINED.
 DR EMBL: 060887; AAC51362.1; JOINED.

DR EMBL: 060888; AAC51362.1; JOINED.
 DR EMBL: 060889; AAC51362.1; JOINED.
 DR EMBL: 060890; AAC51362.1; JOINED.
 DR EMBL: 060891; AAC51362.1; JOINED.
 DR EMBL: 060892; AAC51362.1; JOINED.
 DR EMBL: 060893; AAC51362.1; JOINED.
 DR EMBL: 060894; AAC51362.1; JOINED.
 DR EMBL: 060895; AAC51362.1; JOINED.
 DR EMBL: 060896; AAC51362.1; JOINED.
 DR EMBL: 060897; AAC51362.1; JOINED.
 DR EMBL: 060898; AAC51362.1; JOINED.
 DR EMBL: 068567; AAC50812.1;
 DR EMBL: 005572; AAB03816.1;
 DR MIM: 248500;
 DR INTERPRO: IPR000602;
 DR PFAM: PF01074; Glyco_hydro.38; 1.
 KW Hydrolase; Glycosidase; Glycoprotein; Lysosome; Zymogen; Stnall;
 KW Disease mutation; Polymorphism.
 KM
 ET SIGNAL 1 48
 ET CHAIN 49 1010
 ET CHAIN 49 344
 ET CHAIN 345 428
 ET CHAIN 429 600
 ET CHAIN 601 881
 ET CHAIN 882 1010
 ET CARBOHYD 132 132
 ET CARBOHYD 309 309
 ET CARBOHYD 366 366
 ET CARBOHYD 496 496
 ET CARBOHYD 644 644
 ET CARBOHYD 650 650
 ET CARBOHYD 691 691
 ET CARBOHYD 765 765
 ET CARBOHYD 831 831
 ET CARBOHYD 929 929
 ET CARBOHYD 988 988
 ET VARIANT 71
 ET VARIANT 277 277
 ET VARIANT 311 311
 ET VARIANT 336 336
 ET VARIANT 354 354
 ET VARIANT 355 355
 ET VARIANT 401 401
 ET VARIANT 412 412
 ET VARIANT 713 713
 ET VARIANT 749 749
 ET VARIANT 808 808
 ET CONFLICT 185 185
 ET CONFLICT 342 342
 ET CONFLICT 383 383
 SQ SEQUENCE 1010 AA; 113673 MW; 11425F6A340F33E1 0Kc04;
 Query Match 4.48; Score 94.5; DR 1; Length 1010;
 Best Local Similarity 17.7%; Pred. No. 4.2;
 Matches 71; Conservative 46; Mismatches 134; Indels 151; Gaps 17;
 QY 89 ARHRDDQWVFAIDYDEKPKIPQPVVPSDCFAAAMPSQYALASQAFARALAVANN 147
 DB 11 AKGLDLSACHWIMS-----RAIKRPDLPPI--CPFIILAAACARAGVETCPVGMNIN 64
 QY 148 -NVLRRQUNPKG-----QY-----EKSYPTGRPLKSLAVEMILANIT----- L 184

[illegible]

NO	DOMAIN	2492	2493	CATALYTIC SITE
SCSEQUENCE	2493	AA:	271979	MM: 100A87263015BFB C80644
Q	Query Match	4.4%	Score 94.5;	Id: 1; Length 2493
R	Best Local Similarity	22.6%	Prod. No. 162	
M	Matches 552	Conservative 31;	Mismatches 72;	Indels 85; Gaps 122
OY	46 QVTDKFLIMLNROVMGVAVYNLEERPKOMELAR +			WGA- 85
DB	1948 QLTDSDEVIITANKRLMOWYSYA-----VDIAKQGNLPMIAOKIKDFALSYAE 1990 +			111
OY	86 -----DELAKHCHDIDENNYFA----- +			110E 106
DB	1991 ESTIMWVLSVALITYKSLGKNGSGSLNFASYSKNSIALDKRAIKRFRELLTJRIKLADILDEV 2050 +			113E 106
OY	107 GRPLQPVNVESTPEFAMMI-SOYLASAQEAQKATLGAYNNVLNRQINPRQYKRSYG 166 +			111
DB	2051 APPEDVAIVAFED- LKNSLSLTFNNMO-----PARKHLYTLKRRIIDCVKRTG 2104 +			111
OY	167 TTRFKSLAVPLIANTIEEMELLPTPIVEEVLAQTAVEVMDLDELGLMKREVNTTG 226 +			111
DB	2105 IAWWSE FVSVAL- WVF-TVVQQLD- ELMREILISEEG KEVYMSG 2151 +			113E 114
OY	227 EFV 229 +			111
DB	2152 ELI 2154 +			111
RESULT 12				
HCVH_PANIN				
ID	HCVH_PANIN	STANDARD:	PRF:	657 AA.
AC	P10787:			
DT	01-JUN-1989 (rel. 11, created)			
UT	01-JUN-1989 (rel. 11, last sequence update)			
UL	30-MAY-2000 (rel. 39, last annotation update)			
ODE	HEMOCYANIN B CHAIN.			
OS	Paratubus interruptus (Galitortia spiny lobster).			
OEC	Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;			
EOC	Eumalacostrata; Eucarida; Decapoda; Pleocyemata; Palinur-			
EOC	pallinoidae; Palinuridae; Paratubus.			
RN	111			
RP	SEQUENCE:			
KX	MEDLINE: 89091175.			
KA	Jerkel P.A., Bak H.J., Soeter N.M., Verelken J.M., Beintema J.J.: "Paratubus interruptus hemocyanin. The amino acid sequence of subunit b and anomalous behaviour of subunits a and b on polyacrylamide gel electrophoresis in the presence of SDS."			
KE	Int. J. Biochem. 178:403-412(1988).			
KL	121			
RN	X-RAY CRYSTALLOGRAPHY (4.2 ANGSTROMS).			
RP	Gaykema W.P.J., Hol W.G.J., Verelken J.M., Soeter N.M., Bak H.J., Beintema J.J.:			
KA	"4.2 A structure of the copper-containing, oxygen carrying protein paratubus interruptus haemocyanin."?			
KI	Nature 409:241-29(1984).			
KL	[3]			
RN	X-RAY CRYSTALLOGRAPHY (4.2 ANGSTROMS).			
RP	MEDLINE: 90064489.			
KA	Vollbrecht A., Hol W.G.J.:			
KI	"Crystal structure of hexameric haemocyanin from Paratubus interruptus refined at 4.2 A resolution."			
KL	J. Mol. Biol. 209:249-279(1989).			
OY	-1- FUNCTION: HEMOCYANIN ARE COPPER-CONTAINING OXYGEN CARRIERS COORDINATING FREELY DISSOLVED IN THE HEMOLYMPH OF MANY MOLLUSCS AND ARTHROPODS.			
OY	-1- SIMILARITY: HEXAMER OF A NUMBER OF DIFFERENT CHAINS, OF WHICH A, B, AND C HAVE BEEN IDENTIFIED.			
OY	-1- SUBCELLULAR LOCATION: EXTRACTABLE.			
OY	-1- TISSUE SPECIFICITY: HEMOLYMPH.			
OY	-1- MISCELLANEOUS: THE B CHAIN CONTAINS TWO COPPER BINDING SITES.			
OY	-1- THREE HISTIDINE RESIDUES ARE LIKENS TO EACH COPPER ION.			
OY	-1- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY. HEMOCYANIN			

CC SUBFAMILY.
 DR PIR: S02707: BHI.OB.
 DR HISSP: P04254: IIEY.
 DR INTERPRO: IPR00096: ?
 DR INTERPRO: IPR002227: ?
 DR PFAM: PF00472: hemocyanin_1.
 DR PROSITE: PR00187: HEMOCYANIN.
 DR PROSITE: PS00209: HEMOCYANIN_1.
 DR PROSITE: PS00210: HEMOCYANIN_2.
 DR PROSITE: PS00498: HYDROLYASE_2.
 KW Respiratory protein: oxygen transport: copper: glycoprotein:
 hemocyanin.
 F1 DISULFID 93 98
 F1 DISULFID 483 502
 F1 DISULFID 562 609
 FT METAL 194 194
 FT METAL 198 198
 F1 METAL 224 224
 F1 METAL 444 444
 F1 METAL 448 448
 FT METAL 384 384
 FT CARBOHYD 167 167
 FT VARIANT 42 42
 F1 VARIANT 49 49
 F1 VARIANT 122 122
 F1 VARIANT 130 130
 FT VARIANT 132 132
 FT VARIANT 561 561
 SQ SEQUENCE 657 AA: 75492 MW: D2445E265B70C47E CRC64:

Query Match 4.38: Score 94: DB 1: Length 657:
 Best Local Similarity 21.48: Pred. No. 4.2:
 Matches 60: Conservative 38: Mismatches 96: Indels 86: Gaps 16:
 QY 18 HQVLPFWPKSLDKGCGYFCLPKKQVETPKLFLQNKV--WQFAVYNNLEKRP 75
 DB 138 HQVLPFWPKSLDKGCGYFCLPKKQVETPKLFLQNKV--WQFAVYNNLEKRP 75
 QY 76 Q--WLELR-----HDAFLR---HGDQDQ---NMVF 101
 DB 245 FLMHKKITREGRADLETKRYGGEPVVDNHFEDVAVAHVHDLLETSKRIHADLDCGY 304
 QY 102 ALADPK--PIKQPNV--FSIDFAMAFSCYALASQALALQAVNNVLRQNNK 157
 DB 405 ITDSGHTIDRQKGIELDGLIENSSWVQYGSLENTALIM-----LQSGDQPH 358
 QY 158 QY-----EKSYGTPKLSLAVPMILANI LEMMLPPTVEVLAQIVPMDP 210
 DB 459 GKFNLPVGMDEHFLAIDPSEFLRKHYMDNFKKHLSFPPYIHDL-----EF 408
 QY 211 LDFELGMRKAVPTGE---FVDSFRELN--PKHGLE 244
 DB 409 ---AKVVDGALADGLITFDPEVSYSLAVNGENIE 444

RESULT 13
 GUNC_PSEFL
 ID GUNC_PSEFL STANDARD: PRT: 748 AA.
 AC I27083:
 F1 01-A06-1992 (rel. 24: created)
 F1 01-A06-1992 (rel. 24: last sequence update)
 F1 01-JUN-1994 (rel. 24: last annotation update)
 DE ENDOGUCINASE C PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
 DE (CELLULOSE) (EC) (CELLULOSEINASE C).
 ON ERIC.
 OS Pseudomonas fluorescens.
 OS Bacteria: Proteobacteria: gamma subdivision: Pseudomonadaceae:
 PS Pseudomonas.
 RN 111
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 38 47.
 RC STRAIN SP. CELLULOASA
 RX MEDLINE: 92001996.

RA Pectinase L.M.A., Hazelwood U.K., Barker P.J., Gilbert H.J.
 RI "The cellulohydrolase from pseudomonas fluorescens subsp. cellulosa
 RI consists of multiple functional domains."
 RL Biochem. J. 279:793-799(1991).
 CC -1- CATALYTIC ACTIVITY: ENDOTHELIALSIS OF 1,4 BETA D-GULO-SIDE
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE BINDING DOMAIN
 CC (CBM).
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
 CC HYDROLASES).

CC this SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL contribution
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/annouce/>
 CC or send an email to license@isb.slb.ch).

DR EMBL: X61299: CA43597.1: ?
 DR EIR: S16849: S16849.
 DR EIR: S19652: S19652.
 DR HISSP: P07966: IEXH.
 DR INTERPRO: IPR001547: ?
 DR INTERPRO: IPR001919: ?
 DR INTERPRO: IPR002845: ?
 DR INTERPRO: IPR00554: CBL_2.
 DR PFAM: PF00554: CBL_2.
 DR PFAM: PF02013: CBL_5.
 DR PFAM: PF00150: cellulase_1.
 DR PROSITE: PS00561: CHD_BACTERIAL_1.
 DR PROSITE: PS00659: GLYCOSYL_HYDROL_45.
 KW Cellulose degradation: Hydrolase: Glycosidase: Stenol.
 FT SIGNAL 1 38
 FT CHAIN 39 748
 FT DOMAIN 39 146
 FT DOMAIN 137 179
 FT DOMAIN 228 280
 FT DOMAIN 281 748
 FT DISULFID 39 133
 FT ACT_SITE 503 503
 FT ACT_SITE 653 653
 FT ACT_SITE NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 748 AA: 80197 MW: AC96104137932B76 CRC64:

Query Match 4.38: Score 92.5: DB 1: Length 748:
 Best Local Similarity 18.48: Pred. No. 4.2: Indels 159: Gaps 17:
 Matches 71: Conservative 45: Mismatches 111:

QY 70 RLEPKQMLEIARRGADFLARRHGDQNNVFALDQKDLKQPVNVSIDCFAMAFSCY 129
 DB 432 RLDAKPPYVDANRGYDFTR-----EY 454
 QY 130 ALASGAEAKAIALQAVNNVLRQNNPKGQYKSPTRKSLAVPMILANI----- 182
 DB 455 SCSAINPSSVTRIDAVIK-----QKMIANIELIARISAKLSVSNLIGHVEN 502
 QY 183 -LTEMMLPPTVEEVLAVREVMTDLPDELGLMREAVPTGEVSE----- 242
 DB 503 EPPDYTAENKKNVEAV-QAINEVNPMLLIYEGLISANNTQGTPTDVSPPHRSSTL 561
 QY 233 -----KRIILNPGHGLFAMFM--MLAQR--SDQGLGQ 264
 DB 562 NPNNGENLYEAGNPENIPKDLFLSHITGSPVTFVQGMVAJVECAELREHAAQAR 621
 QY 265 AIAVNLN--LEFVAMDEER-----GIFYPIDRGHPQOLEW----- 607
 DB 622 Q-KIVINIVLPGQWEPHFGYIKELAYGIIA-----QHGNNMIMQAKSSVAFKAW 673
 QY 308 HLETLVALAKGHAQVGGEEKWQ-----WDER--VHGVAMSHFLAHE----- 446
 DB 674 -----SHITINVGQVQVAAASYFKRKGINQWYS-MNPESATIMWYITPWL 721
 QY 347 -----YQEWQYINRQGVILNIRK 366

DR	PDB:	1HC5; 4I-JAN-94.
DR	PDB:	1HC6; 4I-JAN-94.
DR	PDB:	1HCY; 4I-JAN-94.
DR	INTERPRO:	IPRO00896; -1
DR	INTEGRIC:	IPIR01227; -1
DR	PEAM:	PF00372; hemocytin_1.
DR	PRINTS:	PF00197; HEMOCYTIN.
DR	PROSITE:	PS00209; HEMOCYTIN_1.
DR	PROSITE:	PS00210; HEMOCYTIN_2.
DR	PROSITE:	PS00496; YHOSNASE_2.
KW	Respiratory protein; oxygen transport; copper; glycoprotein; alpha-structure.	
KW	DISEFID	93
ET	DISEFID	483
ET	DISEFID	502
F1	DISEFID	562
F1	CAMPOMD	167
F1	METAL	194
ET	METAL	198
ET	METAL	224
ET	METAL	444
F1	METAL	448
F1	METAL	484
F1	TORN	8
ET	HELIX	10
ET	TORN	18
ET	TORN	21
F1	HELIX	28
F1	TORN	39
F1	TORN	43
ET	STRAND	46
ET	TORN	47
ET	TORN	51
F1	HELIX	53
F1	TORN	62
F1	TORN	74
ET	HELIX	77
ET	TORN	92
ET	STRAND	93
F1	HELIX	96
F1	TORN	106
ET	HELIX	112
ET	HELIX	127
ET	HELIX	137
ET	HELIX	142
F1	TORN	145
F1	HELIX	148
F1	TORN	161
ET	STRAND	166
ET	TORN	174
ET	TORN	179
F1	HELIX	182
F1	TORN	186
F1	HELIX	188
ET	TORN	203
ET	HELIX	207
ET	TORN	216
F1	HELIX	218
F1	TORN	237
F1	TORN	248
ET	STRAND	251
ET	STRAND	261
ET	STRAND	267
F1	STRAND	275
F1	STRAND	280
F1	TORN	281
ET	STRAND	284
ET	TORN	286
ET	HELIX	287
F1	STRAND	304
F1	TORN	308
F1	STRAND	312
ET	TORN	318
ET	HELIX	320

Tue Jan 2 15:22:43 2001

us-09-645-321-1.rsp

Page 12

Job Time: 460 sec


```

10 17 4606 PRELIMINARY: PRT: 607 AA.
AC 17 4606:
11 01-FEB-1997 (TREMURIEL, 02, created)
12 01-FEB-1997 (TREMURIEL, 02, last sequence update)
13 01-JUN-2000 (TREMURIEL, 14, last annotation update)
14 HYPOHETICAL: 70.0 KDa PROTEIN.
15 SIK1855.
16 Synchocystis sp. (strain pcc 6803).
17 Bacteria: Cyanobacteria: Chroococcales: Synchocystis.
18 [1]
19 SEQUENCE FROM N.A.
20 MEDLINE: 97061201.
21 KANEKO T., SATO S., KOTANI H., IANAKA A., ASAMIZU E., NAKAMURA Y.,
22 MIYAJIMA N., HIRASAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
23 HOSOUCHI T., MATSUO A., MURAKI A., NAKAZAKI N., NARIO K., OKUMURA S.,
24 SHIMO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
25 TADADA S.
26 "Sequence analysis of the genome of the unicellular cyanobacterium
27 Synchocystis sp. strain pcc6803. 1. Sequence determination of the
28 entire genome and assignment of potential protein-coding regions."
29 J. Mol. Biol. 310:1-16(1996).
30 EMBL: D90908; BAA17650.1; -.
31 Hypothetical protein.
32 SHOEENGE 607 AA; 70003 MW; 873442F37B6A0F3 C0664;
33
34 Query Match 5.1%; Score 108.5; DB 2; Length 607;
35 Best Local Similarity 22.7%; Pred. No. 0.42; Indels 129; Gaps 25;
36 Matches 89; Conservative 47; Mismatches 127;
37
38 58 NQVQVAFVFNKRLPKVLEIARHGAFIARHGR -DQNGN---WYFALLQDQKPLRQ 112
39 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
40 159 SLVYGFALAF -MHTGQVLEAEKGEYELREHREYKEDILYVHALDVQGEKQK 217
41 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
42 114 EV -NFSIE -FAAM -AFSQ -YALASQAEKATALQVNNVLEKQ -IN----- 155
43 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
44 218 FASFGIAYALAYVAYALAG-----PVGTYKQIGIRIIRHADIKLPKPF 269
45 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
46 156 -FKQYKSYKTRPL-----KSLA-----VMIILANTLLEKMLPTT 194
47 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
48 270 LKSEYGFTHSLDPLMLPRSESLGPKARKKNNNSVADHAYVILN-----WL--AT 321
49 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
50 195 VEEVLA-----QVLEFVMTDLPDPIQIMKAVPIGEEV-----DSFGRL 237
51 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
52 422 GEGYKADMLETFDITKEKFPY-----DISPFGDERFEQSDITWQWQ 368
53 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
54 208 N---PRLGLEAMWEMQIAQSGDQLOEOALATLNTLEFAMDEEFGITFFLDK--QG 292
55 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
56 609 NRAVVSNNIKIAMNIMKNNLSKSEKYVYIARKIMAVQSDQDQGMVYVERELIN 428
57 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
58 294 HPR-QULEW DQKLMWULETLVA--LA-----KRIQATNGEKWMQFEVHYVAMSIF 442
59 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
60 429 ISKCHQFVWHKRWQEOALVALLAGITIDEEYHRIQSE----- 471
61 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62 343 ADPEYQWPGYLNKQ---EVLLN---LKG 367
63 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
64 472 ASAEYNAWLELDGATFYNLANIPYLAG 504
65 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
66
67 RESULT 5
68 Q9SHN5 PRELIMINARY: PRT: 662 AA.
69 AC Q9SHN5:
70 01-NOV-1996 (TREMURIEL, 01, created)
71 01-NOV-1996 (TREMURIEL, 01, last sequence update)
72 01-MAY-2000 (TREMURIEL, 13, last annotation update)
73 HEMOCYANIN PROCURSOR [CONTAINS: HEMOCYANIN, LONG FORM, HEMOCYANIN,
74 SHORT FORM].
75 Penaeus vannamei (Penaeid shrimp) (European white shrimp).
76 Eukaryota: Metazoa: Arthropoda: Crustacea: Malacostraca:
77 Eumalacostraca: Eucarida: Decapoda: Penaeoidea: Penaeidae:
78 Penaeidae: Penaeus.
79 [1]

```

```

80 SHOEENGE FROM N.A. AND SEQUENCE OF 15-52.
81 TISSUE=HEPATOYAN/HEAS.
82 MEDLINE: 9709475.
83 RA Seelios P.Y., Lemoine S., Van Wormhout A.:
84 "Molecular cloning of hemocyanin cDNA from Penaeus vannamei
85 (Crustacea, Decapoda): structure, evolution and physiological
86 aspects."
87 J. Exp. Zool. 407:15-158(1997).
88 CC -1- FUNCTION: HEMOCYANIN ARE COPPER CONTAINING OXYGEN CARRIERS
89 ACQUIRING: FREELY DISSOLVED IN THE HEMOLYMPH OF MANY MOLLUSKS AND
90 ARTHROPODS.
91 EMBL: X82502; CAA57880.1; -.
92 HSSP: P04254; I061.
93 DR INTERPRO: IPR000896; -.
94 DR INTERPRO: IPR002227; -.
95 DR PFAM: PF00472; hemocyanin; 1.
96 DR PRINTS: PR00187; HEMOCYANIN.
97 DR PROSITE: PS00209; HEMOCYANIN_1; 1.
98 DR PROSITE: PS00210; HEMOCYANIN_2; 1.
99 DR PROSITE: PS00498; TYROSINASE_2; 1.
100 KW Signal: Respiratory protein; oxygen transport; copper;
101 KW Endoplasmic reticulum; Metal-binding.
102 ET SIGNAL 1 14
103 ET CHAIN 15 662 HEMOCYANIN, LONG FORM.
104 ET PROPEP 15 20
105 ET CHAIN 21 662 HEMOCYANIN, SHORT FORM.
106 ET METAL 208 208 COPPER A (BY SIMILARITY).
107 ET METAL 212 212 COPPER A (BY SIMILARITY).
108 ET METAL 248 248 COPPER B (BY SIMILARITY).
109 ET METAL 358 358 COPPER B (BY SIMILARITY).
110 ET METAL 362 362 COPPER B (BY SIMILARITY).
111 ET METAL 398 398 COPPER B (BY SIMILARITY).
112 SU SEQUENCE 662 AA; 74980 MW; 52B0137557E75178 C0664;
113
114 Query Match 4.8%; Score 104; DB 5; Length 662;
115 Best Local Similarity 22.7%; Pred. No. 1.4;
116 Matches 87; Conservative 47; Mismatches 149; Indels 104; Gaps 21;
117
118 23 PPRKYSILKQGGVFTGIRKGGVETDKEFTWLNQNGQVAF -FYNLEP -KPYQL 78
119 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
120 216 PEW-----WNAVYGHHLQKQ-----ENPFWIHQITVRFARISNYLIVGHEQWN 268
121 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
122 79 E-----IARHGAFIARHG-----RQDGMWYFALQ 105
123 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
124 264 KPIYDAFAPPIFYKYGQGFARFPIQNVKFEVDVARTREWYVESKRIATAIYIT -VDS 322
125 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
126 106 FQKPL---KQYVNVSDCTAAMASQYALASQAEKATALQVNNVLRQHNPQGYE 161
127 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
128 323 EKKHIDISNKKGIDIDIDIDIDIDIDIDIDIDIDIDIDIDIDIDIDIDIDIDID 376
129 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
130 162 KSYPS-----TRPLKSLAVPMILANTLLEKMLPTT-----VEEVAQTYKE 205
131 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
132 377 LP-PRVLEHFFATIRIDISPRIRHKYMNIDFKRHNDIPPIKADIDPSTVSILAVVGE 435
133 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
134 206 VMTDLPDPIQIMKAVPIGEEVDSFGRLINQHGCEAMQEMQIAQSGDQLOEOAL 245
135 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
136 446 LFTYFEPFVSLIN--AVDAQEGITQWELSTYVRLNKEITFTLVENKQAEK----- 487
137 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
138 324 QEKWMQFEVHYVAMSIF--FAPEY 448
139 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
140 533 EKKSTSSVIVDVSDIDIPFAFAAG 599
141 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
142
143 RESULT 6
144 Q9SHN5 PRELIMINARY: PRT: 422 AA.
145 AC Q9SHN5:
146 01-MAY-2000 (TREMURIEL, 14, created)

```


DE SERINE THREONINE PROTEIN KINASE SNF1P.
 GN SNF1.
 GN Cochlodolus carthagen (Bipolaris zeicola).
 GN Cochlodolus carthagen SNF1 gene is required for cell wall degradation
 GN Eukaryota: Fungi: Ascomycota: Pleosporales; Pleosporaceae;
 GN Cochlodolus.
 GN 111
 GN SEQUENCE FROM N.A.
 KA STRAIN SH111.
 KA Fookari N.J., Scott-Craig J.S., Walton J.D.;
 KA "Cochlodolus carthagen SNF1 gene is required for cell wall degradation
 KA enzyme expression and virulence on maize."
 KA Submitted (JUN 1999) to the EMBL/GenBank/DBJ databases.
 KL Submitted (JUN 1999) to the EMBL/GenBank/DBJ databases.
 LR EMBL: AF159253; A003441.1; .
 LR INTERPRO: I0800719; .
 LR INTERPRO: I0802290; .
 LR EFMAM: PF00609; KINASE_1.
 LR PROSITE: PS00107; PROTEIN_KINASE_ATP_1.
 LR PROSITE: PS0108; PROTEIN_KINASE_ST_1.
 LR PROSITE: PS0011; PROTEIN_KINASE_P000; 1.
 KW Kinase.
 SJ SEQUENCE 880 AA; 97949 MW; CDIDGCL24624E16 CRC64;
 Query Match 4.28; Score 90.57; DB 3; Length 880;
 Best Local Similarity 19.18; Pred. No. 25;
 Matches 49; Conservative 35; Mismatches 102; Indels 71; Gaps 9;
 QY 25 WEKSLDQKQVYFTLDKRGVFDITKFLWLNQWQGFVFNKLEPRKQMETLRH; 84
 DB 564 WPKVIRKRAQ-----KSSS-----KSSISQ-----ADPRKRSRHS 618
 QY 85 ADPLARHQRWQWYATLQDQKPLKQPNVPSICFAAMFSAVALASQAFKATALO 144
 DB 619 QDSISHSSEDDG-----SRKGSREPLESVNN-----GTSQELKRGGRKK 661
 QY 145 AVNNVLRQGNPKQYKSYQGTFRKSLAVPMLANIILPMKLIPTIVEVLAQIVR 204
 DB 662 HYNH-----NWDVHYVEDVWVIAHPL-----KQMPFGVAHNSSTHSSR 704
 QY 205 EVMIEFLPELIMEAVPTGFEVDFEGRLLNPGHTEAMWPMQIAQKSDRLQEQ 264
 DB 705 ---VIAHNSQARRKRSSTNS-----TSSACHGVGMIIISKAGSVSEIHVNIDE 752
 QY 265 AIAVNIILLYAMDEE 281
 DB 753 AVYIMSTQSYDQDF 769
 RESUL 15
 Q59910 PRELIMINARY: PRF: 417 AA.
 A5 059910;
 DT 01-NOV-1996 (FEBM01.01, last sequence update)
 DI 01-NOV-1996 (FEBM01.01, last sequence update)
 DI 01-MAY-2000 (FEBM01.14, last annotation update)
 DI CYTOCHROME P450 (EC 1.14.14.1).
 OS Streptomyces filadelf.
 OS Streptomyces filadelf.
 OS Bacteria: Filimicrobia: Actinobacteria: Actinobacteridae;
 OS Actinomycetales: Streptomycetales; Streptomyces.
 KA SEQUENCE FROM N.A.
 KA STRAIN-159245;
 KA MEDLINE: 95075319.
 KA Molsion Davies L.A., Gundliffe E.;
 KA "Analysis of liver tylosin biosynthetic genes from the tylosin region of
 KA the Streptomyces fradiae genome."
 KA Mol. Microbiol. 13:349-355(1994).
 GN -1 SIMILARITY: RELINGS TO THE CYTOCHROME P450 FAMILY.
 DE EMBL: D08223; AA021341.1; .
 DE HSSP: G00441; 10XA.
 DE INTERPRO: IPR001128; .
 DE INTERPRO: IPR002497; .
 DE PFAM: PF00067; Pf50; 1.

DR PRINS: PR00459; HP450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN 1.
 KW oxidoreductase; Monooxygenase; Electron Transport; Membrane; Heme-
 FT BINDING 466 466 HEME (BY SIMILARITY).
 SJ SEQUENCE 417 AA; 47224 MW; A76194224DFCEH1 CRC64;
 Query Match 4.28; Score 90; DB 2; Length 417;
 Best Local Similarity 21.28; Pred. No. 9.4;
 Matches 77; Conservative 56; Mismatches 111; Indels 140; Gaps 20;
 QY 55 WLNQWQWQ-----FAVYNKLEKQWFIARHGADE 148H30 94
 DB 50 WDSKQWQVFRVDYDLVSTINQLEPSDSQVFPVPELALIMQGTFGQIDPRH; 107
 QY 95 QGNWYFALDGEKPLKQPNVSPCFAMAFSQAVALASQAFKATALOANNVIRKH 154
 DB 108 -----PLRK-----LVSAAPPRKATLEPRATIPQALDGL 140
 QY 155 NPGQVEKSYNGTRPKSLAVNMLA-----NLTMMRM-----LP-PIVFEV 198
 DB 141 KEGQIDVSDLAFLPVIVIAELLSIPAEPRDLEKRWVVIINNGMEYFNIPDH SET 200
 QY 199 LAQIVKRWMLDPLDPLQIMREAVPTGFEVSP-----EGR-----IINNG 240
 DB 201 MCPALKE-WQWLYRRLAKRE--TLDIDMSGLDAEVGRRLDEELVNIVALIAG 257
 QY 241 HIDEA -----WIMMD-----IQKSDRLQEQALAVNIINLEYWDEERKITYFL 288
 DB 258 HISSATILQNLPLVLDHREDAELRADRIIDGATE- EILRYK -SPNNIRPLKE 312
 QY 289 --PROGHPPQV-----LEMDQKIMVHLFETVALAKHQAIGQKQWQFRRVHWAMSH 342
 DB 313 DTLILSRPKAGQWVAV-----IASANDSK-----HIF 341
 QY 343 ADPE 346
 DB 342 SDPD 345

Search completed: January 1, 2001, 22:57:47
 Job time: 452 sec

[illegible]

ALLEGMENTS

[illegible]

KILPATRICK

Factorial 12: Yarn-diameter (4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14; 15; 16; 17; 18; 19; 20; 21; 22; 23; 24; 25; 26; 27; 28; 29; 30; 31; 32; 33; 34; 35; 36; 37; 38; 39; 40; 41; 42; 43; 44; 45; 46; 47; 48; 49; 50; 51; 52; 53; 54; 55; 56; 57; 58; 59; 60; 61; 62; 63; 64; 65; 66; 67; 68; 69; 70; 71; 72; 73; 74; 75; 76; 77; 78; 79; 80; 81; 82; 83; 84; 85; 86; 87; 88; 89; 90; 91; 92; 93; 94; 95; 96; 97; 98; 99; 100; 101; 102; 103; 104; 105; 106; 107; 108; 109; 110; 111; 112; 113; 114; 115; 116; 117; 118; 119; 120; 121; 122; 123; 124; 125; 126; 127; 128; 129; 130; 131; 132; 133; 134; 135; 136; 137; 138; 139; 140; 141; 142; 143; 144; 145; 146; 147; 148; 149; 150; 151; 152; 153; 154; 155; 156; 157; 158; 159; 160; 161; 162; 163; 164; 165; 166; 167; 168; 169; 170; 171; 172; 173; 174; 175; 176; 177; 178; 179; 180; 181; 182; 183; 184; 185; 186; 187; 188; 189; 190; 191; 192; 193; 194; 195; 196; 197; 198; 199; 200; 201; 202; 203; 204; 205; 206; 207; 208; 209; 210; 211; 212; 213; 214; 215; 216; 217; 218; 219; 220; 221; 222; 223; 224; 225; 226; 227; 228; 229; 230; 231; 232; 233; 234; 235; 236; 237; 238; 239; 240; 241; 242; 243; 244; 245; 246; 247; 248; 249; 250; 251; 252; 253; 254; 255; 256; 257; 258; 259; 260; 261; 262; 263; 264; 265; 266; 267; 268; 269; 270; 271; 272; 273; 274; 275; 276; 277; 278; 279; 280; 281; 282; 283; 284; 285; 286; 287; 288; 289; 290; 291; 292; 293; 294; 295; 296; 297; 298; 299; 300; 301; 302; 303; 304; 305; 306; 307; 308; 309; 310; 311; 312; 313; 314; 315; 316; 317; 318; 319; 320; 321; 322; 323; 324; 325; 326; 327; 328; 329; 330; 331; 332; 333; 334; 335; 336; 337; 338; 339; 340; 341; 342; 343; 344; 345; 346; 347; 348; 349; 350; 351; 352; 353; 354; 355; 356; 357; 358; 359; 360; 361; 362; 363; 364; 365; 366; 367; 368; 369; 370; 371; 372; 373; 374; 375; 376; 377; 378; 379; 380; 381; 382; 383; 384; 385; 386; 387; 388; 389; 390; 391; 392; 393; 394; 395; 396; 397; 398; 399; 400; 401; 402; 403; 404; 405; 406; 407; 408; 409; 410; 411; 412; 413; 414; 415; 416; 417; 418; 419; 420; 421; 422; 423; 424; 425; 426; 427; 428; 429; 430; 431; 432; 433; 434; 435; 436; 437; 438; 439; 440; 441; 442; 443; 444; 445; 446; 447; 448; 449; 450; 451; 452; 453; 454; 455; 456; 457; 458; 459; 460; 461; 462; 463; 464; 465; 466; 467; 468; 469; 470; 471; 472; 473; 474; 475; 476; 477; 478; 479; 480; 481; 482; 483; 484; 485; 486; 487; 488; 489; 490; 491; 492; 493; 494; 495; 496; 497; 498; 499; 500; 501; 502; 503; 504; 505; 506; 507; 508; 509; 510; 511; 512; 513; 514; 515; 516; 517; 518; 519; 520; 521; 522; 523; 524; 525; 526; 527; 528; 529; 530; 531; 532; 533; 534; 535; 536; 537; 538; 539; 540; 541; 542; 543; 544; 545; 546; 547; 548; 549; 550; 551; 552; 553; 554; 555; 556; 557; 558; 559; 560; 561; 562; 563; 564; 565; 566; 567; 568; 569; 570; 571; 572; 573; 574; 575; 576; 577; 578; 579; 580; 581; 582; 583; 584; 585; 586; 587; 588; 589; 590; 591; 592; 593; 594; 595; 596; 597; 598; 599; 600; 601; 602; 603; 604; 605; 606; 607; 608; 609; 610; 611; 612; 613; 614; 615; 616; 617; 618; 619; 620; 621; 622; 623; 624; 625; 626; 627; 628; 629; 630; 631; 632; 633; 634; 635; 636; 637; 638; 639; 640; 641; 642; 643; 644; 645; 646; 647; 648; 649; 650; 651; 652; 653; 654; 655; 656; 657; 658; 659; 660; 661; 662; 663; 664; 665; 666; 667; 668; 669; 670; 671; 672; 673; 674; 675; 676; 677; 678; 679; 680; 681; 682; 683; 684; 685; 686; 687; 688; 689; 690; 691; 692; 693; 694; 695; 696; 697; 698; 699; 700; 701; 702; 703; 704; 705; 706; 707; 708; 709; 710; 711; 712; 713; 714; 715; 716; 717; 718; 719; 720; 721; 722; 723; 724; 725; 726; 727; 728; 729; 730; 731; 732; 733; 734; 735; 736; 737; 738; 739; 740; 741; 742; 743; 744; 745; 746; 747; 748; 749; 750; 751; 752; 753; 754; 755; 756; 757; 758; 759; 760; 761; 762; 763; 764; 765; 766; 767; 768; 769; 770; 771; 772; 773; 774; 775; 776; 777; 778; 779; 780; 781; 782; 783; 784; 785; 786; 787; 788; 789; 790; 791; 792; 793; 794; 795; 796; 797; 798; 799; 800; 801; 802; 803; 804; 805; 806; 807; 808; 809; 810; 811; 812; 813; 814; 815; 816; 817; 818; 819; 820; 821; 822; 823; 824; 825; 826; 827; 828; 829; 830; 831; 832; 833; 834; 835; 836; 837; 838; 839; 840;

[illegible]

DB 921 TGGTAACTTAAATGGAAATGAAATATGAAGCTGGTGGGCAACACCTGAAAGTATGATT 980
 QY 940 GCGCTAAGCAAGTACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 999
 DB 941 GGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1040
 QY 1000 GCGCTAAGCAAGTACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1059
 DB 1041 GCGCTAAGCAAGTACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1100
 QY 1060 GCGCTAAGCAAGTACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1119
 DB 1101 GCGCTAAGCAAGTACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1160
 QY 1120 GCGCTAAGCAAGTACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1157
 DB 1161 GCGCTAAGCAAGTACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1198

RESULT 3
 HOMOLOGY 1500 bp mRNA 981 12-SEP-1998
 DEFINITION Homo sapiens mRNA for retin-binding protein, complete cds.
 ACCESSION D10242.1 GI:220052
 VERSION D10242.1 GI:220052
 KEYWORDS retin-binding protein,
 Homo sapiens (strain:Caucasian) 3-month old male Wilms' tumor,
 cell_line:G-401 (ATCC CRL 1441) cDNA to mRNA, clone:lambda HK66.

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 1300)

REFERENCE
 AUTHORS Iwano, H., Takahashi, S., Fokui, K. and Miyake, Y.
 TITLE Genetic and molecular properties of human and rat retin-binding
 proteins with reference to the function of the leucine zipper motif
 JOURNAL J. Biochem. 110 (4), 493-500 (1991)
 MEDLINE 9210649
 COMMENT Data kindly submitted in computer-readable form by: Hiroyasu Inoue
 Department of Biochemistry
 National Cardiovascular Center Research Institute
 5-7-1 Fujishiro-dai
 Suita Osaka 565
 Japan
 Feature: 06-844-5012 x2458.

FEATURES
 Source location/Qualifiers
 1..1300

/organism="Homo sapiens"
 /strain="Caucasian"
 /db_xref="taxon:9606"
 /cell_line="G-401 (ATCC CRL 1441)"
 /cell_type="Wilms' tumor"
 /clone="lambda HK66"
 /dev_stage="3-month old"
 /sex="male"
 41..1284
 /from_start=1
 /product="Retin-binding protein"
 /protein_id="AA01082.1"
 /db_xref="GI:220052"

CHS

POLYA_signal

POLYA_site 1300 372 C 444 G 243 T
 HASE_COUNI 241 A
 ORIGIN

Query Match 11.7% Score 147.4 DB 84 Length 1400
 Best local similarity 48.4% Pred. No: 45-302
 Matches 542 Conservative 0 Mismatches 556 Indels 21 Gaps 5
 QY 60 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 119
 DB 93 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 152
 QY 120 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 179
 DB 153 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 212
 QY 180 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 243
 DB 213 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 272
 QY 244 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 287
 DB 273 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 332
 QY 288 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 347
 DB 343 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 392
 QY 348 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 404
 DB 493 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 452
 QY 405 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 464
 DB 453 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 512
 QY 455 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 524
 DB 513 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 572
 QY 525 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 584
 DB 573 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 632
 QY 585 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 641
 DB 633 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 692
 QY 642 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 701
 DB 693 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 752
 QY 702 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 761
 DB 753 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 812
 QY 762 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 818
 DB 813 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 872
 QY 819 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 878
 DB 873 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 932
 QY 879 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 938
 DB 933 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 992
 QY 939 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 998
 DB 993 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1052
 QY 999 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1058
 DB 1053 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1112
 QY 1059 GGTATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1118

106	544	CTGATGCT
-----	-----	---

[illegible]

[illegible]

*	87117	87216:	gap of	100 bp	in length
*	87217	88245:	contig of	1029 bp	in length
*	88246	88345:	gap of	100 bp	in length
*	88346	89493:	contig of	1148 bp	in length
*	89494	89593:	gap of	100 bp	in length
*	89594	90596:	contig of	1003 bp	in length
*	90597	90659:	gap of	100 bp	in length
*	90697	93350:	contig of	2654 bp	in length
*	93351	93450:	gap of	100 bp	in length
*	93451	94536:	contig of	1086 bp	in length
*	94537	94636:	gap of	100 bp	in length
*	94637	95797:	contig of	1161 bp	in length
*	95798	95897:	gap of	100 bp	in length
*	95898	97063:	contig of	1166 bp	in length
*	97064	97163:	gap of	100 bp	in length
*	97164	98234:	contig of	1071 bp	in length
*	98235	98334:	gap of	100 bp	in length
*	98335	99650:	contig of	1316 bp	in length
*	99651	99750:	gap of	100 bp	in length
*	99751	101345:	contig of	1595 bp	in length
*	101346	101445:	gap of	100 bp	in length
*	101446	107114:	contig of	5669 bp	in length
*	107115	107214:	gap of	100 bp	in length
*	107215	110997:	contig of	3783 bp	in length
*	110998	111097:	gap of	100 bp	in length
*	111098	114007:	contig of	2910 bp	in length
*	114008	114107:	gap of	100 bp	in length
*	114108	115183:	contig of	1076 bp	in length
*	115184	115283:	gap of	100 bp	in length
*	115284	118707:	contig of	4424 bp	in length
*	118708	118807:	gap of	100 bp	in length
*	118808	120082:	contig of	1275 bp	in length
*	120083	120182:	gap of	100 bp	in length
*	120183	122094:	contig of	1912 bp	in length
*	122095	122194:	gap of	100 bp	in length
*	122195	123403:	contig of	1209 bp	in length
*	123404	123503:	gap of	100 bp	in length
*	123504	127244:	contig of	3741 bp	in length
*	127445	127344:	gap of	100 bp	in length
*	127345	128970:	contig of	1626 bp	in length
*	128971	129070:	gap of	100 bp	in length
*	129071	130490:	contig of	1420 bp	in length
*	130491	130590:	gap of	100 bp	in length
*	130591	132146:	contig of	1556 bp	in length
*	132147	132246:	gap of	100 bp	in length
*	132247	137509:	contig of	5263 bp	in length
*	137510	137609:	gap of	100 bp	in length
*	137610	138922:	contig of	1313 bp	in length
*	138923	139022:	gap of	100 bp	in length
*	139023	141189:	contig of	2167 bp	in length
*	141190	141289:	gap of	100 bp	in length
*	141290	142374:	contig of	1085 bp	in length
*	142375	142474:	gap of	100 bp	in length
*	142475	144249:	contig of	1775 bp	in length
*	144250	144349:	gap of	100 bp	in length
*	144350	147088:	contig of	2739 bp	in length
*	147089	147188:	gap of	100 bp	in length
*	147189	148659:	contig of	1471 bp	in length
Localize/Qualifiers					
1. 148659 "Homo sapiens"					
/db_xref="taxon:9606"					
/chromosome="13"					
/clone="RP11-128P23"					
/clone_lib="RP11-11.1"					
1. 2308					
/note="Assembly:fragment:00336"					
fragment_chain:1"					
misc_feature					
409... 4575					
/note="assembly:fragment:01005"					
fragment_chain:1"					
misc_feature					
4676... 77354					

```

fragment_chain:1
7455..8720
/notes="assembly_treatment:00376
treatment_chain:1"
8821..10650
/notes="assembly_fragment:00514
fragment_chain:2"
10751..12192

misc_feature
Query Match 3.4%; Score 39.8; DB 84; Length 148659;
Best Local Similarity 57.7%; Pred. No. 1.9;
Matches 71; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 1030 gacccaagtatgggaatgattgctgaactgaatgcgcgaggaagatgaactgaac 1089
||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110214 GTTCGGAGACATGGTGAATGTTCAAGGGTGTGTTGTGTGGTGGACAGAGGTGCAAGATC 110274
||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1090 cttaaaaggaagaatgaatgaagatgctgcttcacgtaccccgagctctatgactctggc 1149
||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110274 CAAGTAGCTGGAGATACCCACAGAGCAAGCCCAAGTCCTCCACATCATGATGATTAACCTGG 110334
||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1150 gaa 1152
||
Db 110334 AAA 110336

RESULT 11
AL163539
LOCUS Homo sapiens chromosome 13 clone RP11-444G16, *** SEQUENCING IN
DEFINITION PROGRESS ***, 40 unordered pieces.
ACCESSION AL163539
VERSION AL163539.4 GI:9230998
KEYWORDS HTS: HTGS_PUANEL.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Mammalia; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 154112)
Barton,T.
Direct Submission
Submitted (09-JUN-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonequests@sanger.ac.uk
On Jul 16, 2000 this sequence version replaced 41:8573850.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA34616
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 131709 bases at least Q40
Consensus quality: 130219 bases at least Q30
Consensus quality: 142824 bases at least Q20
Insert size: 150212; sum-of-configs
Insert size: 159639; agarose-fp
Quality coverage: 2.85x in Q20 bases; sum-of-configs quality
coverage: 2.68x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 40 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
3426: contig of 3426 bp in length

```

```

* 4427 4526: gap of 100 bp
* 4527 6166: count iq of 2640 bp in length
* 6167 6266: gap of 100 bp
* 6267 12956: count iq of 6696 bp in length
* 12957 14056: gap of 100 bp
* 14057 48191: count iq of 35145 bp in length
* 48192 48291: gap of 100 bp
* 48292 58146: count iq of 9845 bp in length
* 58147 68286: gap of 100 bp
* 68287 60162: count iq of 1926 bp in length
* 60163 60262: gap of 100 bp
* 60263 62224: count iq of 1962 bp in length
* 62225 62324: gap of 100 bp
* 62325 64498: count iq of 2174 bp in length
* 64499 64598: gap of 100 bp
* 64599 70055: count iq of 5457 bp in length
* 70056 70155: gap of 100 bp
* 70156 74584: count iq of 4429 bp in length
* 74585 74684: gap of 100 bp
* 74685 77947: count iq of 3263 bp in length
* 77948 78047: gap of 100 bp
* 78048 79781: count iq of 1744 bp in length
* 79782 79881: gap of 100 bp
* 79882 80969: count iq of 1088 bp in length
* 80970 81069: gap of 100 bp
* 81070 84623: count iq of 3554 bp in length
* 84624 84723: gap of 100 bp
* 84724 85940: count iq of 1207 bp in length
* 85941 86040: gap of 100 bp
* 86041 87122: count iq of 1092 bp in length
* 87123 87222: gap of 100 bp
* 87223 89098: count iq of 1876 bp in length
* 89099 89198: gap of 100 bp
* 89199 91963: count iq of 2747 bp in length
* 91964 92063: gap of 100 bp
* 92064 93351: count iq of 1316 bp in length
* 93352 93451: gap of 100 bp
* 93452 94651: count iq of 1200 bp in length
* 94652 94751: gap of 100 bp
* 94752 95763: count iq of 1012 bp in length
* 95764 95863: gap of 100 bp
* 95864 97009: count iq of 1146 bp in length
* 97010 97109: gap of 100 bp
* 97110 99457: count iq of 2348 bp in length
* 99458 99557: gap of 100 bp
* 99558 102566: count iq of 3009 bp in length
* 102567 102666: gap of 100 bp
* 102667 109403: count iq of 6647 bp in length
* 109404 109403: gap of 100 bp
* 109404 110703: count iq of 1300 bp in length
* 110704 110803: gap of 100 bp
* 110804 113476: count iq of 2574 bp in length
* 113477 114476: gap of 100 bp
* 114477 118046: count iq of 4570 bp in length
* 118047 118146: gap of 100 bp
* 118147 119856: count iq of 1640 bp in length
* 119857 119956: gap of 100 bp
* 119957 123719: count iq of 3784 bp in length
* 123720 123819: gap of 100 bp
* 123820 128041: count iq of 4222 bp in length
* 128042 128141: gap of 100 bp
* 128142 129496: count iq of 1255 bp in length
* 129497 129496: gap of 100 bp
* 129497 131959: count iq of 2464 bp in length
* 131960 132059: gap of 100 bp
* 132060 138747: count iq of 6678 bp in length
* 138748 138847: gap of 100 bp
* 138848 141491: count iq of 2554 bp in length
* 141492 141491: gap of 100 bp
* 141492 142943: count iq of 1452 bp in length
* 142944 143043: gap of 100 bp
* 143044 147504: count iq of 4461 bp in length
* 147505 147604: gap of 100 bp

```

FEATURES

SOURCE

```

* 147605 149060: count iq of 1456 bp in length
* 149061 149160: gap of 100 bp
* 149161 151795: count iq of 2645 bp in length
* 151796 151895: gap of 100 bp
* 151896 154112: count iq of 2217 bp in length
Location/Qualifiers
1. 154112
/organism "Homo sapiens"
/chr "X"
/chromosome "13"
/cytome "RP11_344516"
/cytome.lib "RP11.2"
1. 3426
/feature "assembly_fragment:00164"
/fragment_chain:1"
4527. 6166
/feature "assembly_fragment:00424"
/fragment_chain:1
c1c0cc_cnd:599
vector_side:left"
6267. 12956
/feature "assembly_fragment:01120"
/fragment_chain:1"
14057. 48191
/feature "assembly_fragment:01176"
/fragment_chain:2"
48292. 58146
/feature "assembly_fragment:00176"
/fragment_chain:2"
58237. 60162
/feature "assembly_fragment:00615"
/fragment_chain:2"
60263. 62224
/feature "assembly_fragment:00044"
/fragment_chain:3"
62325. 64498
/feature "assembly_fragment:00845"
/fragment_chain:3"
64599. 70055
/feature "assembly_fragment:00116"
/fragment_chain:4"
70156. 74584
/feature "assembly_fragment:00675"
/fragment_chain:4"
74659. 77947
/feature "assembly_fragment:00925"
/fragment_chain:5"
78048. 79781
/feature "assembly_fragment:00898"
/fragment_chain:5"
79882. 80969
/feature "assembly_fragment:01043"
/fragment_chain:6"
81070. 84623
/feature "assembly_fragment:00206"
/fragment_chain:6"
84724. 85940
/feature "assembly_fragment:01162"
/fragment_chain:7"
86041. 87122
/feature "assembly_fragment:01249"
/fragment_chain:7"
87223. 89098
/feature "assembly_fragment:01247"
/fragment_chain:8"
89199. 91963
/feature "assembly_fragment:00474"
/fragment_chain:8"
92064. 93351
/feature "assembly_fragment:00094"
/feature "assembly_fragment:00114"
94752. 95763

```

misc_feature

misc_feature	/note="assembly_traitment:00185"
	95864..97009
misc_feature	/note="assembly_fragment:00203"
	97110..99457
misc_feature	/note="assembly_fragment:00300"
	99558..102566
misc_feature	/note="assembly_traitment:00464"
	102667..109303
	/note="assembly_fragment:00479"
Query Match:	4.48; Score 39.8; Dh 64; Length 154112;
Best Local Similarity	57.7%; Pred.No.1.9;
Matches	71; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
CY 1030	gacccaataatgaagaattgttgctagcactgaatccgcggaggagatgttacctaac 1089 + + + + + + + + + + + + + + +
Dh 122946	GTTCGCGACATGGTGAATTCTCAGCGTGCTCTGTGGCGGCCGCAGAGTCGCAAGACC 122995 + + + + + + + + + + + + + + +
CY 1090	ctaaagaaaqaaaatagaaqqagtttcaccagagcccccaactcatggtcttctgcc 1149 + + + + + + + + + + + + + + +
Dh 122996	GAACTAGCTGGGAATACCCAGACGAGAAGCCAGATCCCACATCATCTAATCACTGG 123055 + + + + + + + + + + + + + + +
CY 1150	gaq 1152 + + + + +
Dh 123056	NNA 123058
RESULT 12	
CELLOL.	
LOCUS	2232 bp DNA INV 04-JUN-1993
DEFINITION	Caenorhabditis elegans collagen gene.
ACCESSION	M80650
VERSION	M80650.1 GI:156249
KEYWORDS	collagen.
SOURCE	Caenorhabditis elegans (strain NZ).
ORGANISM	Caenorhabditis elegans (strain NZ) DNA. Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdillida; Rhabdilloidea; Rhabdillidae; Peloterinae; Caenorhabditis.
REFERENCE	1 (sites) CoX,G.N., Fields,C.A., Kramarz,J.M., Rosenzweig,H. and Hirt,D. Sequence comparisons of developmentally regulated collagen genes of Caenorhabditis elegans Gene 76, 331-344 (1989)
JOURNAL	89326131
REFERENCE	2 (bases 1 to 2232) Hirt,D.M. Sequence comparison of the Caenorhabditis elegans dpy-13 and col-34 genes, and their deduced collagen products Gene 120, 261-266 (1992)
AUTHORS	93013043
TITLE	
MEDLINE	
FEATURES	Location/Qualifiers 1..2232 /organism="Caenorhabditis elegans" /strain="NZ" /db_xref="taxon:6239" 381..393 533..539 /gene="alpha-collagen" join(585..665,720..1362,1411..1583) /gene="alpha-collagen" join(585..665,720..1362,1411..1583) /gene="alpha-collagen" /codon_start=1 /product="alpha-collagen" /protein_id="AAA27985.1" /db_xref="GI:156250" /translation="MDETRIKAVFNAVSVAVVAVTVCVTLPEWYVVIHKVR TMNEITFCISKARDIMVEVIALKSPNSNTARQVNDIAVTGGAGSGSCSTLE GPDPACTPGCKPRPKRGAGALGNDRPPQQCEPIETPPMKRPQGPGPQPPE TPHGSGEPGKLPGQIAADNPGRPKRGKPCAGATFHGESDFCRGHPLDGHED PHGAGPGHPGSGPGPCAGDSPPGPKRGKNQPDQOIGACNNGCATGPAQTPTGK KGLTKRYALIDGVFFEDGTR" 666..719
INITIATION	

exon

720..1362

/gene-"alpha-collagen"

/number-1

/partial

/gene-"alpha-collagen"

/number-2

/label-exon2

1463..1410

/gene-"alpha-collagen"

/number-2

1603..1608

/gene-"alpha-collagen"

BASE COUNT 639 a 551 c 534 g 508 t

ORIGIN

Query Match

Best Local Similarity 49.38; Pred. No. 2.6;

Matches 101; Conservative 0; Mismatches 104; Indels 0; Gaps 0

QY 772 gatgcacattataggaqaaagcatttcagtgtatgaaacccttggaatattgccttgg 831

Db 1309 GATGCGAAGGCACCCAGAGATCCCCAGCAGCAACGAGAGAGAGAGAGATATCATATAAGA 1308

QY 832 qatgaagaattttgatgcataatttatcttccttgatcccaaggacaccttcctaatga 891

Db 1369 AATAGATTATAGTGTACTATTTTTAAACTGTGCTTCACAGCAACGCCAGCCAAAGC 1428

QY 892 ctggaatggacaaccaagctctgtgtgcatcttggaacccttgttccttagcgaag 951

Db 1429 GATCAATATGACGATGATGGTGACCCAGAGAGCGAGATGGAATGCTTAGAGATAGAGATCAG 1408

QY 952 ggcacccaagcactgcccagaaa 976

Db 1489 CCGGACCACCCAGAGATCCCAGCAGCA 1513

RESULT 13

CeLF36A4/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted by:

Submitted (01-APR-1996) Robert Waterston

* consists of 98 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1	658:	contig of 658 bp in length
659	1295:	gap of unknown length
1296	2007:	gap of unknown length
2008	2721:	gap of unknown length
2722	3459:	gap of unknown length
3460	4086:	gap of unknown length
4087	4801:	gap of unknown length
4802	5446:	gap of unknown length
5447	6376:	gap of unknown length
6377	6747:	gap of unknown length
6748	7373:	gap of unknown length
7374	7989:	gap of unknown length
7990	9108:	gap of unknown length
9109	9727:	gap of unknown length
9728	10439:	gap of unknown length
10440	11447:	gap of unknown length
11448	12247:	gap of unknown length
12248	13091:	gap of unknown length
13092	13788:	gap of unknown length
13789	14602:	gap of unknown length
14603	15418:	gap of unknown length
15419	15966:	gap of unknown length
15967	16948:	gap of unknown length
16949	17661:	gap of unknown length
17662	18453:	gap of unknown length
18454	19243:	gap of unknown length
19244	19998:	gap of unknown length
19999	20862:	gap of unknown length
20863	22034:	gap of unknown length
22035	23296:	gap of unknown length
23297	24124:	gap of unknown length
24125	25153:	gap of unknown length
25154	25990:	gap of unknown length

* 25991

26855:	contig of 865 bp in length	
26856	27763:	gap of unknown length
27764	28489:	gap of unknown length
28490	29172:	gap of unknown length
29173	29922:	gap of unknown length
29923	30634:	gap of unknown length
30635	31461:	gap of unknown length
31462	32123:	gap of unknown length
32124	32982:	gap of unknown length
32983	33746:	gap of unknown length
33747	34563:	gap of unknown length
34564	35272:	gap of unknown length
35273	36366:	gap of unknown length
36367	37458:	gap of unknown length
37459	38609:	gap of unknown length
38610	39416:	gap of unknown length
39417	40019:	gap of unknown length
40020	41351:	gap of unknown length
41352	43045:	gap of unknown length
43046	44892:	gap of unknown length
44893	45469:	gap of unknown length
45470	46762:	gap of unknown length
46763	47850:	gap of unknown length
47851	49437:	gap of unknown length
49438	50702:	gap of unknown length
50703	51982:	gap of unknown length
51983	53151:	gap of unknown length
53152	53993:	gap of unknown length
53994	55057:	gap of unknown length
55058	56117:	gap of unknown length
56118	57311:	gap of unknown length
57312	58302:	gap of unknown length
58303	59260:	gap of unknown length
59261	60675:	gap of unknown length
60676	61873:	gap of unknown length
61874	63206:	gap of unknown length
63207	64718:	gap of unknown length

GenCore version 4.5
Copyright (c) 1994 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: January 1, 2001, 22:23:50 : Search time 75.08 seconds
(without alignments)
430,487 Million cell updates/sec

Title: US-09-645-321-1

Percent score: 2147
Sequence: 1 MIAHRQELAQYYQALHDP.....CFIVPRALMLCAETLQPV 391

Scoring table:
Gap: 10.0, Gapext: 0.5

Search: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post processing: Minimum Match: 0%
Maximum Match: 100%
Listing first 45 summaries

Database: PIR-55:
1: PIR1:
2: PIR2:
3: PIR3:
4: PIR4:

Field No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2147	100.0	391	2	renin-binding prot
2	694	32.3	402	2	renin-binding prot
3	668	41.1	419	2	renin-binding prot
4	645	30.0	417	2	renin-binding prot
5	148	6.3	418	2	hypothetical 48k p
6	108.5	5.1	607	2	hypothetical 48k p
7	103	4.8	662	2	homocystin procus
8	98.5	4.5	689	2	conserved hypotet
9	97.5	4.5	1007	3	alpha-mannosidase
10	95	4.3	876	2	probable val's prot
11	94.5	4.3	266	2	hypothetical prote
12	94.5	4.4	2493	2	adenylate cyclase
13	93	4.3	657	1	hemocytin chain b
14	92.5	4.3	748	2	cellodextrinase c
15	92	4.3	1067	2	isoleucyl-tRNA syn
16	91.5	4.3	692	2	conserved hypotet
17	91	4.2	391	2	hypothetical prote
18	91	4.2	657	1	hemocytin chain a
19	91	4.2	961	2	alpha-mannosidase
20	90	4.2	417	1	cystochrome p450 ty
21	89.5	4.2	583	2	general alpha-trehal
22	88.5	4.1	698	2	general sporulatio
23	88.5	4.1	803	2	hypothetical prote
24	88	4.1	900	2	alanine-tRNA ligas
25	87	4.1	348	2	glycosyl hydrolase
26	87	4.1	720	2	heat shock protein
27	86.5	4.0	642	2	conserved hypotet
28	86	4.0	1066	2	isoleucine-tRNA 1
29	85.5	4.0	316	2	hypothetical prote

30	85.5	4.0	686	1	ALDYAT	amylase A (EC 3.2
31	85.5	4.0	912	2	D72644	hypothetical prote
32	85.5	4.0	205	2	108164	dynactin alpha heavy
33	85	4.0	339	1	MOHPCP	N-acetylaminoglyc
34	85	4.0	700	2	S57194	calpain (p3, 4, 22
35	85	4.0	833	2	S62136	CD44 protein - cy
36	84.5	3.9	180	2	A32246	70k anti IgM - chi
37	84.5	3.9	536	2	A81710	fibronectin protein
38	84.5	3.9	580	2	D72093	fibronectin protein
39	84.5	3.9	748	2	S41050	fibroblast growth
40	84.5	3.9	750	2	S41051	fibroblast growth
41	84	3.9	465	2	D69785	beta glucosidase b
42	84	3.9	2222	1	A36028	DNA-directed RNA p
43	83.5	3.9	454	2	G64737	polyomavirus late ad
44	83.5	3.9	470	1	NM1VM8	co-alpha-stalidas
45	83.5	3.9	513	2	S48981	hypothetical prote

ALIGNMENTS

RESULT 1
S75649
renin-binding protein-related protein - Synchocystis sp. (strain PCC 6804)
N/A: Alternate names: protein SL1975
C: Species: Synchocystis sp.
A: Variety: PCC 6804
C: Date: 25-Apr-1997 #sequence_revision 25 Apr 1997 #text change 20-Jun-2000
C: Accession: S75649
R: Kaneko, T.; Saito, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, O.; K. Okumura, S.; Shimpou, S.; Takeuchi, G.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-116, 1996
A: Title: Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis

A: Reference number: S74322; MIMD:97061201
A: Accession: S75649
A: Status: nucleic acid sequence not shown; Translation not shown
A: Molecule type: DNA
A: Residues: 1-391 - KAN.
A: Cross-references: EMBL:090912; GIB:AB001349; NITE:053428; PIR:66A32101; U00419
A: Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996.

Query Match	100.0%	Score 2147	DB 2	Length 391
Best local similarity	100.0%	pred. No. 2	16	176
Matches 391	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MIAHRQELAQYYQALHDPVLPFWERKSLDKSGSYETGLDKRGVETDKEITWQNRQ	60	
DB	1	MIAHRQELAQYYQALHDPVLPFWERKSLDKSGSYETGLDKRGVETDKEITWQNRQ	60	
QY	61	VMQFAVYNNKLPKQWMLFARRGADPLAAGKQDQNNMFALDQKRLKQYNNVSYD	120	
DB	61	VMQFAVYNNKLPKQWMLFARRGADPLAAGKQDQNNMFALDQKRLKQYNNVSYD	120	
QY	121	FAAMASQVALASQAEAKAIALQAVNNVLRQHNKRGQYKSYPTIRPKSLAVNPILA	180	
DB	121	FAAMASQVALASQAEAKAIALQAVNNVLRQHNKRGQYKSYPTIRPKSLAVNPILA	180	
QY	181	NLTLEMWMLPFTVEEVLAAQIVREWTDLPDPIGIMREAVTPTGEVDSFEKRLNPG	240	
DB	181	NLTLEMWMLPFTVEEVLAAQIVREWTDLPDPIGIMREAVTPTGEVDSFEKRLNPG	240	
QY	241	HGTEAMFPMQDLQKSLDKQLODQALAVVNIITFADDERFGGIPYLDQGHVQVQILW	300	
DB	241	HGTEAMFPMQDLQKSLDKQLODQALAVVNIITFADDERFGGIPYLDQGHVQVQILW	300	
QY	301	DKLIMVHLEETLVALAKGHQATQOEKQWQEFKSHVYAMSHLPDEGEMFGYLNKRGV	360	
DB	301	DKLIMVHLEETLVALAKGHQATQOEKQWQEFKSHVYAMSHLPDEGEMFGYLNKRGV	360	
QY	361	LNLKAAKMKGCFIVPRALMLCAETLQPV 391		
DB	361	LNLKAAKMKGCFIVPRALMLCAETLQPV 391		

62 WQFVITYNRELE -PKVWLEIARIH:ADFLAKIR- IQDNNWTFALDQEKPLQRYNWF 117
 118 63 WMYRIRYRTEREPRHAOLIIIAAKADDEFLIRKAVAI-IGKKCAVILLRDRCPYKVVORII 122
 118 118 SPCFAAMAF-SYVAIASQJAKAKAIATLQAVNNVIR-RQHINKQGYEKSYVQIRPKSIAPV 176
 124 124 SEFTYIMAMNLMKRAFQYRKYQTEAVEEMQJYHVVQEDASGRLPOLQCAFAAEFRAVP 182
 177 177 MIJANI-LEHMLLPLIIVFEVLAQIVREV-----MIDFIDPEIGIMREAVIPIDEEV 229
 194 194 MMLINIVQI-----GIAPEELAKQYAEIIGVWARRILQHVQDQJAVIFENSPQSKEL 236
 240 240 TSEBELLNPHGIEIEMVPMMDIAJRSIDRLQLEQAI-AVVLNILEYAMTEERGGITYEL 288
 247 247 PGTJRKQNNHLLFAQMFELIRKQDPIKRAHVITKELLIFHSQMGIDHOGIIFYPQ 296
 269 269 DQGHPIPGVITWIKKIWWVHLEIIVALKQHQALQDQK*WQMFIRKVDYAMSHADPEYG 348
 277 277 IAINP-PIQLFWARKLWMPISSEAMIAFLMYSQSQDEVLRLIYQVAYEITFQGFQDEYQ 356
 449 449 EMPEYMRKREIVINIKKKWKQCFEYIRALWICAEII 486
 457 457 EMPEYMRKREIVINIKKKWKQCFEYIRALWICAEII 494
 RESULT 5
 540824
 Hypothetical 48K protein (glna-tdhe region) - Escherichia coli
 GiDate: 19-May-1994 #sequence_revision 01-Sep-1995 #text_change 04-Mar-2000
 GiSpecies: Escherichia coli
 CDate: 19-May-1994 #sequence_revision 01-Sep-1995 #text_change 04-Mar-2000
 GiAccession: S40824; 56193
 RefLink: 111, G.; Burland, V.; Daniels, D.L.; Hattner, F.K.
 Nucleic Acids Res. 21, 3391-3398, 1993
 ATitle: Analysis of the Escherichia coli genome. 111. DNA sequence of the region from 8
 AReference number: S40802; M01D:93347969
 AAccession: S40824
 AStatus: preliminary; nucleic acid sequence not shown; translation not shown
 AModel type: DNA
 AResidues: 1-418 - PILE
 ACross-references: EMBL:J119201; NID:q404961; PIDN:AA03013.1; PID:q304984
 ANote: The nucleotide sequence was submitted to the EMBL Data Library, October 1993
 ARef: Rose, D.J.; Man, B.; Shao, Y.
 Science 272, 1454-1456, 1997
 ATitle: The complete genome sequence of Escherichia coli K-12.
 AReference number: A04720; M01D:97426617
 AAccession: 965194
 AStatus: preliminary; nucleic acid sequence not shown; translation not shown
 AModel type: DNA
 AResidues: 1-418 - PILE
 ACross-references: GB:AB000464; GB:000096; NID:q2467324; PIDN:AA013442.1; PID:q1790313
 AExperimental source: strain K-12, substrain MG1655
 AGenes: ynf
 ASuperfamily: Escherichia coli hypothetical 48K protein (glna-tdhe region)
 Query Match 6.98; Score 148; Dh 2; Length 418;
 Best Local Similarity 21.14; Pred. No. 5.5e-05;
 Matches 82; Conservative 63; Mismatches 195; Indels 78; Gaps 17;
 48 PCTIDKQV FDIKFLIMQNRQVQFAVYNLEIKPQWLEIARHCAPIARHG---- 92
 14 14 PCTIDKQV FDIKFLIMQNRQVQFAVYNLEIKPQWLEIARHCAPIARHG---- 92
 38 PCTIDKQV FDIKFLIMQNRQVQFAVYNLEIKPQWLEIARHCAPIARHG---- 92
 93 93 PCTIDKQV FDIKFLIMQNRQVQFAVYNLEIKPQWLEIARHCAPIARHG---- 141
 88 NCALDKKQVAVAVVNDQVAVDASKQYQHF---FALDAAS-AVILGHEAKILDYI 143
 142 142 NCALDKKQVAVAVVNDQVAVDASKQYQHF---FALDAAS-AVILGHEAKILDYI 143
 144 144 NCALDKKQVAVAVVNDQVAVDASKQYQHF---FALDAAS-AVILGHEAKILDYI 143

187 187 EMLPPTVEEVIAQVIRE-----VMIDFLPELSMKR- AVTPGEEVDSFESRLNNG 240
 197 197 KMDRAIIVASVI IHDVAANNHRYVNEHFDIQMNLIIYNNKINIAHRR-KAQGG---IIG 252
 241 241 HGTAEAMTMDIAQSRGK-----QLEQALAVVINTLEYAM-DEERGITYFELRQSH 294
 253 253 HWTGWRMELIHDALEADQEQFPAWLEDAKINERATKRAADGAGIYVITWEEK 312
 294 294 PPOULFMDQKLMVH-PTIVALKQHQALQDQK*WQMFIRKVDYAMSHADPEYQ 354
 314 314 PVVR -EKVWEIFEAMKIVAVITVIGQYETIWTWYETIKYIMQENSMQJE 368
 354 354 INRGVLLNLKSGKRR - GGFHVR 377
 459 459 IIAQNNVITKVMGKQDIIYHILHIVITR 397
 RESULT 6
 577092
 Hypothetical protein sif185c - Synecocystis sp. (strain PCC 6804)
 GiSpecies: Synecocystis sp.
 AVarity: PCC 6804
 CDate: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
 GiAccession: S77092
 RefLink: 111, S.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O.; Kikukawa, S.; Shimizu, S.; Takemura, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yan
 DNA Res. 3, 109-116, 1996
 ATitle: Sequence analysis of the genome of the unicellular cyanobacterium Syn. 6804
 AReference number: S74322; M01D:97061201
 AAccession: S77092
 AStatus: nucleic acid sequence not shown; translation not shown
 AModel type: DNA
 AResidues: 1-607 - KAN-
 ACross-references: EMBL:D90098; GB:AB001339; NID:q1652725; PIDN:BA017650.1; PID:q165
 ANote: The nucleotide sequence was submitted to the EMBL Data Library, June 1996
 ASuperfamily: Synecocystis hypothetical protein sif185c
 Query Match 5.18; Score 108.5; Dh 2; Length 607;
 Best Local Similarity 22.78; Pred. No. 0.22;
 Matches 89; Conservative 47; Mismatches 127; Indels 129; Gaps 25;
 58 NROVQFAVYNLEIKPQWLEIARHCAPIARHG---IQDNNWTFALDQEKPLQRYNWF 117
 159 SKLVYGRALAF-MDTGDDPILAAEKGTETIRHMRPVKDEIDIIYWHATDVQGRKQK 217
 113 PY-NVPSIDFAM-AESQ-YALASIAQEAALATLQAVNNVLRQ-IRN----- 155
 218 IFASRGIDYDAIPAYQIYALAG-----PIQIVKICIGIKRIIHAQVILKIDKFF 269
 156 -IKQYKSYVIGIRP-----KSLA-----VMILAN-LEHMLLPLI 194
 270 LKSEYGVFSLDPLMDLPRSESLGPNKARKMNSVGDHAFAYLNL-----WL--AT 421
 195 VEVVIA-----QIVREVMDLPEIGIMREAVIPIDEEV-----DSFGKRL 237
 322 GQKVDMLPYTDTIRKYPVQ-----DHSFQDERPYEMASHDITWQWQ 368
 238 N---PHTGFAEMFMMLIORSQDROLQEQALAVVINTLEYAMDEEFGSIFELDR--QG 292
 369 NNAVVGINILKIANIMHNSIKSEKYYVHAKKIALIMAVQSDQKQWYIVVIRILAN 428
 429 HSRCHQFVWIRKAWQEQEALAYLILMLDDEEYHRRQE----- 471
 443 ADFEYGHVGYINRKG- FVILN-----KGG 367
 472 ASAFYAMVFLIDIRGTYFNVAIANGIPIYAGG 503

RESULT 7

homologous protein - peptidyl shimp (bacterial strain)

Accession: EMBL:U026185

Created: 15 Jul 1995 #sequence revision 10 Nov 1996 #text change 19 May 2000

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

Accession: S55487

QY 56 LQNKVQWQFAVYNNLDEKPKQM-----LEIARHQADELARHQKQ- DQNWY----- 101
 DB 120 VEI-----LAFSTR-----WMHQUTNATGEVVD-----LVQGRLETFANSGVMNNEA 162
 QY 102 -----ALUCLDKPLKQYNNF--SITFAAMAFNSYALASQULAKALAUAYNNVL--RR 152
 DB 163 AHHYATLHACMLDLEPFDLEKQKQRYVAMHIDPQSHQVASTPAQMDPDLFFQRL 222
 QY 153 QUNRQGYEKYPTREKLKSLAVNMLLANTLEMFW-----LPPTYVEVLAQTVHEVM 207
 DB 223 DQGRKRYVE-----NLDG EYVWRASASIKRYA-----ALFISVL 258
 QY 208 IDPELPELQIKREAVPIGEVDSFESRIENHQHQTAMWYEMDIAQSHDROUQQAIA 267
 DB 259 ENIYNPEKIDWDLTGADKPEV---EQRSEPEYNAEELVNFQZLATAVDQHPRTNITM 315
 QY 268 VNLITLKYAMIDEEGCIYYELDRQCHPYQULDWQKLMWHLFLLVALAKG-HQAIDQE- 325
 DB 316 IMDSDFQYE-----NANMFRRNIDRLQIYNADQANDSRV 351
 QY 326 ---KQWMEFEVDYAMS-----HFAHFEYQGEVLYNRR 357
 DB 352 NLYSTPAVYWEINKAMNIMSVKQDHPYPAKHQHWSTYPSR 397

RESULT 10

probable vds protein - Mycobacterium tuberculosis (strain H37RV)

C/Spectrum: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C/Accession: G70863

C/Ref: S.T. Broese, K. Parkhill, J. Garnier, T. Churcher, C. Harris, D. Gordon, S. Connor, R. Davies, K. Devlin, K. Fellwell, T. Gillies, S. Hamilton, N. Holtroyd, S. Rajandream, M.A. Roberts, J. Rutter, S. Seeger, K. Skellern, S. Squares, S.

Nature 393, 547-544, 1998

Authors: Squires, K.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.

All the participating the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; MUID:98295987

A/Accession: G70863

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1 876 (90%)

A/Cross-references: db:AL021246; db:AL124356; NID:4261507; PIDD:CAA16025.1; PID:q279148

A/Experimental source: strain H37RV

C/Genetics:

Antigen: vds

C/Superfamily: valine--trna ligase

Query Match 4.4% Score 95; DB 2; Length 876;

Best Local Similarity 19.8%; Pred. No. 5.3;

Matches 74; Conservative 47; Mismatches 141; Indels 122; Gaps 18;

QY 25 WEKYSIDQKQNAVYPTLDRKQVFD--DKEFLMION-----RQWQFAVYNNLEP 73
 DB 117 WKES--GATAGQMRKLDLDGWSRQRETMENLSKAVATEKRLVYIALLY-RAER 171
 QY 74 KQMLEIARHQADELARHQKQW-----YFALDQEKPLKQYNNVSTCFAMATSYQY 129
 DB 172 LVNWSVYQIA-SLEFVNRIVECHVSFRKSYLDLS-----QPHI VVAITRVELMDQI 226
 QY 130 ALASQULAKALAUAYNNVLKQHNFRQNGYSEFGTRPLKSLAVNMLLNTLEMF 189
 DB 227 ALAV-----HFDERKRLVGT--SLAHPIVDELA----- 256
 QY 190 IDPELPELQIKREAVPIGEVDSFESRIENHQHQTAMWYEMDIAQSHDROUQQAIA 267
 DB 257 -----VANEHVDPEFGAVKVIHAID-FINDELGVGRHQLPMSIILDK 299
 QY 244 GGLLNKGRITTEAMFMMLIAQSGIRQULQDQALAVNLTLEYAMDEEGGCIYYELDRQCH 293
 DB 400 GRVVDIGIRFIDM-----DRFARVAVKQALAAQGVVEKKD-----YIHSWGH 344

QY 294 PPOUE-WQDKL--WVHLELV-----ALAKG--HQAIDQERWQWTFRHHYA 338
 DB 345 SERSGEIEERKLSQWNVKVESLAKKAGDQVNRQDLVHIAHSEPPWTSNVLDMHWW 404
 QY 339 ----WSHFAHFEY 348
 DB 405 KQLWCHRIPIWQ 418

RESULT 11

hypothetical protein F13A7.10 - Caenorhabditis elegans

C/Spectrum: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C/Accession: T20806

C/Ref: Murray, A.

Submitted to the EMBL data library, March 1997

A/Reference number: Z19327

A/Accession: T20806

A/Status: preliminary; translated from GB/EMBL/GenBank

A/Residues: 1 266 (MIL)

A/Cross-references: EMBL:Z93477; PIDD:CA807572.1; GSDPB:G000024; PEST:F13A7.10

A/Experimental source: clone F13A7

C/Genetics:

Antigen: CESP:F13A7.10

A/Map position: 5

A/Introns: 27/1; 182/1

Query Match 4.4% Score 93.5; DB 2; Length 266;

Best Local Similarity 20.4%; Pred. No. 1.4;

Matches 50; Conservative 33; Mismatches 96; Indels 67; Gaps 9;

QY 109 PLKQYVNSICFAAMAFNSYALASQULAKALAUAYNNVLKQHNFRQNGYSEFGTR 168
 DB 34 ISNTHFVSMRQDAQKALVAA--VKKIUSEVIN--IMQKPIWYFKNYKQGV 87
 QY 169 P-----LKSIAVEMILANTLEMFLLPTTVEEVLQAQVYVITFLPELST 217
 DB 88 PLIEHACQKQKIVESAVIDPYIDILPEEK-TLSDPYEKVQDKILDFISQILHAFSR 146
 QY 218 MRKAVTPIGEVDSFESRIENHQHQTAMWYEMDIAQSHDROUQQAIAVNIILEFAM 277
 DB 147 VTRAIKMPDELKEFES-----LTKATEEA 171
 QY 278 DEEPQGITV-----FIDRQGHIPVQLEMDQKIWVH-LEIIVAIARHQADELARHQK 328
 DB 172 ESILGAFVNSISQGVQVLLVYS-----FQRYWMLFELHPLPDSNFQNGYPKLS 226
 QY 329 QWFEKV 334
 DB 227 QWFKAI 232

RESULT 12

A55481

adenylate cyclase (Ec 4.6.1.1) uae1 - smut fungus (Ustilago maydis)

C/Spectrum: Ustilago maydis (corn smut)

C/Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 09-Sep-1997

C/Accession: A55481

C/Ref: R. Duncan, G. Barrett, K. Kronstad, J.

Genes Dev. 8, 2805-2816, 1994

A/Title: CAMP regulates morphogenesis in the fungal pathogen Ustilago maydis

A/Reference number: A55481

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1 2493 (90%)

A/Cross-references: db:U39018; NID:q603939; PIDD:q603940

A/Note: nucleotide sequence not given; amino acid sequence not complete

C/Genetics:

